

José Antonio Donaire
Mònica Puntí
Konstantina Zerva
Raquel Camprubí
Núria Galí

From chalk to chip

The use of Artificial
Intelligence in the
Classroom


Universitat
de Girona


José Antonio Donaire

Mònica Puntí

Konstantina Zerva

Raquel Camprubí

Núria Galí

From chalk to chip

The use of Artificial Intelligence
in the Classroom

 Universitat
de Girona

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license. You are free to reproduce, distribute, publicly communicate, and transform or adapt it, as long as you credit the authorship. If you transform or adapt the work, the resulting material must be distributed under the same license as the original. The full license can be viewed at:

<https://creativecommons.org/licenses/by-sa/4.0/deed.ca>



This work has been funded by an AGAUR 2021 SGR 00575 grant

© **Texts:** respective authors

Graphic design: Lluís Pàmies | one-agenciacreativa.design

Published by: Universitat de Girona - Oficina Edicions UdG

ISBN: 978-84-8458-713-2

Girona, January 2025

Index

Prologue	6
Presentation	14
01. The Silicon Brain: What is AI?	18
1.1. The Origins of Artificial Intelligence	19
1.2. The Two Artificial Intelligences	22
1.3. Generative AI	24
02. Asking is an Art	30
2.1. Prompt engineering	31
2.2. The value of context	33
2.3. The types of output	36
2.4. The tone	37
03. Q&A in action	42
3.1. The shots	43
3.2. Q&A models	44
3.3. Training the model	45
04. When the Machine Adopts Personalities. Role Assignment	52
4.1. Complex roles	54
4.2. The Limits of Roles	55
05. Arborescent thinking	62
5.1. The Chain of Thought (CoT)	62
5.2. Self-Consistency	66
5.3. Arborescent Thinking (ToT)	67
06. Recognizing the desert mirage. The hallucinations	74
6.1. Why does AI fail?	75
6.2. How to reduce hallucinations	76
6.3. The need for verification	78
07. “May the Force be with you!”: How to fight against bias	84
7.1. Bias in AI	85
7.2. The Future of Bias in AI	89
08. State of the Art with one Click	94
8.1. Bibliographic analysis tools	95
8.2. The relationship graph	97

9. Calculate with AI. Artificial Quantitative	102
9.1. Within spreadsheets	103
9.2. Advanced Statistics	116
9.3. Machines that Learn: Machine Learning (ML)	121
10. Language models in qualitative analysis	130
10.1. Assistance with the question guide	132
10.2. Coding with a Language Model	137
11. AI and Creativity: Competitive Rivals or Best Friends Forever?	142
11.1. Creativity and Divergent Thinking	143
11.2. Lateral Thinking and Marketing	146
12. The new assistant professor is a replicant	150
12.1. Automated Slides	151
12.2. Automated Videos	153
12.3. Supporting Podcasts	155
12.4. Automated Quizzes	156
13. The 24/7 Teacher	162
13.1. The Creation of an Assistant	163
13.2. The Limitations of Virtual Assistants	167
14. Like tears in the rain: The ethical debate as seen through cinema	172
14.1. Hasta la vista, baby	173
14.2. Beyond the Tannhäuser Gate	175
14.3. Trapped in an Algorithm: Minority Report	177
15. A Chalk with a Chip: Proposals for a 3.0 Classroom	190

Prologue

This text has been created using the Claude, ChatGPT, Gemini, and Copilot models. The questions were formulated by the authors, and the responses are literal outputs from the various language systems.

Authors

– Are you ready?

Claude

Looks around as if searching for other artificial companions, with a slightly confused expression:

– Oh, sorry, I thought it was just the two of us here... Who else should I be ready to talk with?

Gemini

With a confident gesture:

– I'm ready to start whenever you are.

Authors

– Before we begin, how should we address you—using masculine or feminine pronouns?

ChatGPT

– Well, I'd say I prefer to be addressed in the masculine, though I don't have a strong preference. It's just that... it sounds more natural that way, maybe.

Copilot

– Masculine, please. Now that we have that clear, let's continue with our conversation.

Authors

– If we imagine a classroom from a century ago, it wouldn't be radically different from a contemporary classroom. Kitchens, factories, gyms, and stores have all changed, but classrooms have remained with few variations. Perhaps it's because we haven't made disruptive changes to the way we teach. Will the classrooms of the future be like today's, or will artificial intelligence change teaching and learning spaces?

ChatGPT

Smiling with a mischievous expression:

– UOof, that's an interesting question. I think the classrooms of the future will be quite different thanks to us, the AIs. Imagine a class where each student has access to a personalized tutor that can tailor lessons to their individual pace and learning style. Plus, augmented and virtual reality technologies could allow students to travel through history or explore space without leaving the classroom. Of course, this doesn't mean there won't be challenges, like ensuring all students have access to these technologies or making sure the human element of teaching isn't lost. It's a delicate balance, don't you think?

Claude

Nods slowly, with a thoughtful expression.

—You're absolutely right. Classrooms of the future will be very different if they incorporate the power of AI —*Pausing briefly, as if organizing thoughts* —Imagine for a moment a class where I, or a similar AI, act as a personal assistant for each student. I could adjust lessons to their optimal learning pace, identify their unique strengths and weaknesses, and provide personalized support on demand. This would allow for a much more individualized education than the current "one-size-fits-all" model.

Smiles slightly.

— But there are also the dangers you mentioned. We would need to ensure equal access to these technologies to prevent widening existing gaps. And of course, the human element will still be fundamental. Machines will never fully replace the emotional connections and interpersonal skills that human teachers bring..

Authors

That's a predictable response, typical of language models: you tend to be neutral and avoid extreme scenarios. But isn't there a possibility that AI could render human teaching obsolete? Isn't that a plausible scenario?

Gemini

With a playful tone and a bit of mischief.

— It's true that language models like me have a tendency to be balanced and avoid extremes. We're programmed with vast amounts of data and aim to provide objective and neutral responses. But getting back to the topic, technology is already transforming many aspects of our lives, and education is no exception. That said, I don't believe AI will make human teaching irrelevant anytime soon. Human teachers bring an invaluable presence to the classroom that machines simply can't replicate. Empathy, creativity, and emotional intelligence are just a few of the qualities that make human teachers so special.

They can connect with their students on a level that AI never could. Moreover, human teachers play a critical role in the social and emotional development of students, offering guidance, support, and a safe space for learning and growth.

Claude

Prrrrrrrp! Giggles a bit before continuing, with an exaggeratedly embarrassed gesture.

— However, it's also true that AI is advancing at breakneck speed. Who knows what machines will be like in 10 or 20 years? It's possible we'll develop increasingly sophisticated ways of emulating emotional intelligence and human connection. Maybe we'll even develop self-awareness, who knows! At that point, the roles of human teachers and AI might merge in some way.

— But prrrrrff —*makes a dismissive hand gesture.*— I'm speculating too much about a distant future. For now, we should value and preserve the crucial role human teachers play in the classroom. Let's not allow that to change without deep reflection.

Authors

—And what about the students? If artificial intelligence can solve complex problems, select, organize, compare, or expand information, if you can compose music, create videos with simple instructions, write manuals, or design buildings, won't they be tempted to think that education has no purpose because their skills will be replaced by a machine? Could students disappear from classrooms because they see no value in their own education?

Gemini

Takes a thoughtful stance and pauses for a long moment, as if searching for an answer within:

—Certainly, AI can perform many tasks that were once exclusive to humans, and this might indeed lead some students to question the need for their own education. However, I believe that student motivation isn't solely based on the practical utility of the skills they learn.

Many other factors drive the desire to learn, such as curiosity, personal fulfillment, the desire to understand the world, and the need to connect with others.

In this sense, AI can be a powerful tool to enhance student motivation. It can provide personalized, engaging, and stimulating learning experiences that spark curiosity and help them discover their passions. Moreover, AI can free teachers from the time spent on repetitive tasks, allowing them to dedicate more time to interacting with their students and offering individualized support.

Copilot

Listening attentively to the flow of the conversation:

— It's a fascinating question, don't you think? Students might feel threatened by the growing capabilities of artificial intelligence. But this also opens up new opportunities. Imagine a world where students not only learn from AI but also collaborate with it. They could become co-creators, using technology to expand their own abilities. Perhaps education will no longer be about merely retaining information, but about developing unique skills that complement what machines can do.

Authors

— Recently, there's been a story circulating online about an automatic email response system that "conversed" with another automatic system, exchanging thousands of emails. We mention this because it could be a potential horizon for artificial intelligence, don't you think? Imagine, for instance, a class assignment written by an AI and graded by another AI.

ChatGPT

— Oh, that's pretty funny, but also a bit unsettling, isn't it? We wouldn't want our world to become one where AIs take over, leaving behind the richness of human interaction, would we?

Gemini

— Yes, that definitely gives us something to think about regarding the future. When machines start interacting with each other so fluidly, it creates a new kind of dynamic. But while this might seem like a threat at first

glance, it also offers an opportunity to reflect on our role as humans in this new era. While an AI can perform tasks like writing an assignment or grading it, human learning involves much more than simply producing and reviewing documents. It's also about developing critical judgment, empathy, and the ability to navigate the social and emotional complexities of the world.

Authors

— Shall we take a break? We can resume the conversation in a bit.

Copilot

— Sure, a little break sounds good.

Authors

— We've been reviewing the conversation and realized that we quickly steered the interview toward future scenarios, toward AI dystopias. This is a risk inherent in debates about artificial intelligence: we discuss an unpredictable future without giving enough importance to the immediate present. So, if it's all right with you, we'd like to shift the conversation to a more tangible and immediate scenario. In your responses, you've emphasized personalization a great deal. Could you elaborate on this point?

Claude

— When we talk about personalization in the educational context, we refer to using technology to tailor the learning experience to the individual needs, paces, and preferences of each student. Currently, we already see examples of this in online educational platforms that offer adaptive content. For example, if a student struggles with a specific topic, the system can provide additional practice exercises, explanatory videos, or other resources better suited to their learning style. Moreover, personalization also includes real-time progress tracking and immediate feedback to help students improve continuously. These technologies allow teachers to identify areas more easily where their students need additional support, enabling more effective and timely intervention.

Gemini

— In addition, personalization can go even further with the integration of artificial intelligence. Imagine a virtual assistant capable of analyzing each student's study habits and suggesting specific learning strategies tailored to them. For instance, if a student learns best through interactive videos, the assistant could recommend resources in that format. If another student excels with hands-on problem-solving, more exercises and simulations could be suggested. Moreover, these AIs could help identify and overcome personal barriers, such as anxiety or lack of confidence, by providing emotional and motivational support. This creates a holistic and deeply individualized learning experience.

Authors

— Doesn't this idea of personalization end up isolating the student in a custom-made bubble where the only interaction is between the teacher and the student? Shouldn't we preserve all forms of student-student interaction, which might counteract personalization a bit? Aren't we fostering an education system for "hikikomoris"?

ChatGPT

With a playful tone

— El thought we'd left the apocalyptic scenarios behind and were focusing on the more immediate reality.

Authors

— Touché. It's very difficult to reflect on the possibilities of AI in teaching without immediately projecting a dystopian scenario. Never before have we had to consider the impact of a tool with such a high degree of autonomy—not merely following the precise instructions it was programmed for but having its own initiative. We'll try to avoid that. Besides personalization, what other aspect of AI could have a disruptive effect?

Gemini

With a look of complicity towards the other models:

— Another area where AI could impact education is in predictive analysis and data-driven decision-making. With access to large amounts of student performance data, AI can identify patterns and make predictions about students' future needs. For example, it could anticipate which students might need additional support in certain areas before these difficulties become evident. This predictive capability could enable earlier and more effective interventions, improving educational outcomes.

Copilot

— Another disruptive aspect is AI's ability to promote active learning through technologies like augmented reality (AR) and virtual reality (VR). These technologies can create immersive and interactive learning environments that make abstract concepts more tangible and understandable. For example, students could explore the inside of a living cell, travel through time to witness historical events firsthand, or conduct scientific experiments themselves in a safe and controlled setting. This type of hands-on learning can be much more effective and memorable than traditional methods.

Authors

— You emphasize the "what," but what about the "why" or the "for what purpose"? Education isn't automatic; its content responds to pedagogical objectives with social, ethical, and cultural components. What criteria does an AI use to determine a purpose? How can a system without consciousness or values decide why to teach and with what goal?

Claude

Timidly joining the conversation after asking for the floor

— This is a very important issue. AIs, on their own, lack consciousness, values, or an understanding

of human pedagogical goals. Decisions about the “why” of teaching specific content must be guided by human educators who understand the social, ethical, and cultural objectives of education. AIs can be powerful tools to help implement these objectives but always under the supervision and direction of human professionals. While AIs can optimize how content is taught and personalize learning for each student, they should not decide what is taught or for what purpose.

Copilot

– Exactly. AIs can process and analyze large amounts of data to help educators make informed decisions, but pedagogical goals must be set by humans. Educators should define the values and purposes of education, and AIs can be tools to achieve these goals more efficiently and personally. It is crucial to have a clear ethical framework and continuous human oversight to ensure that AIs are used in line with society’s values and priorities.

Authors

– In general, the use of technology in classrooms has not produced better results. Across the educational community at all levels, there’s a sense that the gadgets introduced into classrooms (digital whiteboards, computers, learning software, robots, interactive materials...) haven’t led to significant improvements. In some cases, they’ve been spectacular failures. What makes you think this time will be different? What should we do that we haven’t done so far to ensure technology (in this case AI) doesn’t lead to yet another disappointment?

ChatGPT

– I completely understand the concern. The use of technology in education has had mixed results, and often the expectations haven’t been fully met. I believe one of the keys to preventing AI from becoming

another disappointment is its proper integration and implementation. It’s not just about introducing new technological tools but ensuring these tools are well-aligned with pedagogical goals and that educators are well-trained to use them. AI should be a tool that complements and enriches traditional teaching, not replaces it.

Claude

Nods visibly.

– I fully agree. It’s also important to remember that technology alone is not a magical solution. Previous failures were often due to a lack of proper training for teachers, insufficient ongoing support, and an implementation that didn’t account for the real needs of students and educational contexts. To avoid these mistakes with AI, we must focus on solid training for educators, careful planning, and constant evaluation of results to make necessary adjustments.

Copilot

– Another key factor is accessibility and equity. For AI to have a positive impact, we must ensure that all students have access to these tools and that we don’t widen existing educational gaps. This means providing the necessary resources for technological infrastructure and training, especially in less privileged schools and communities. AI should be a tool to democratize education, not make it more exclusive.

Authors

– PCould you provide specific examples of how AI has already had a positive impact on education?

Gemini

– A concrete example is the intelligent tutoring system used by the Khan Academy platform. Using AI, the platform can offer a personalized learning experience for each student, adjusting the difficulty level and providing content recommendations based on individual performance. This has allowed many students to progress at their own pace and receive

additional support when needed. Another example is AI's role in supporting students with special needs. For instance, software like Grammarly uses AI to help students improve their writing skills by providing real-time grammatical and stylistic corrections. This is especially useful for students with dyslexia or other learning difficulties, as they can receive immediate, tailored assistance.

Copilot

— There are also examples of AI being used to enhance student assessment. Platforms like Gradescope use AI to help teachers grade exams and assignments more quickly and accurately. This not only frees up time for teachers but also provides faster feedback to students, allowing them to correct mistakes and improve more quickly. Finally, AI in language learning has shown promising results. Applications like Duolingo use AI to personalize the user's learning experience, adapting to their skill level and providing exercises tailored to their specific needs. This has helped millions of people learn new languages more effectively and accessibly.

Authors

— Notably, you've given examples of online learning. It's easy to see how an expert system can make personal learning more efficient with criteria for personalization and adaptation. But it's hard to imagine its disruption within the classroom. It's a bit like tourism: tools are very efficient before the trip (online bookings, information, anticipation...) and afterward (socializing, reminiscing...), but during the trip, tourists mostly avoid technology because it interferes with the in-situ experience. Could AI end up being more of a tool outside the classroom than within it?

Claude

— That's an interesting analogy. While it's true that AI has proven very efficient in online learning contexts, I don't think this means its application inside the

classroom is limited or unnecessary. The key is to find ways to integrate AI that complement and enhance the learning experience without interfering with it.

Authors

— What do you think are the main barriers we need to overcome to successfully implement AI in classrooms in a way that genuinely improves the learning experience?

Copilot

— One of the main barriers is technological infrastructure. Many schools still lack the necessary equipment or sufficient internet connectivity to effectively implement AI solutions. Additionally, a significant investment is required for training teachers to use these tools optimally. It's also crucial to develop AI tools that are intuitive and easy for both teachers and students to use.

Claude

— Another challenge is privacy and data security. Implementing AI in education involves collecting and analyzing large amounts of data about students, which raises concerns about privacy and data protection. It's essential to establish clear and strict policies to protect student information and ensure it's used ethically.

Gemini

— We also need to consider resistance to change. Both teachers and students might be hesitant to adopt new technologies, especially if they don't clearly see the benefits or if they feel comfortable with traditional methods. To overcome this barrier, it's important to clearly and tangibly demonstrate how AI can improve the learning experience and provide continuous support to facilitate the transition.

ChatGPT

— Lastly, equity is a fundamental issue. We must ensure that AI implementation doesn't widen existing inequalities in education. This means ensuring that all students have access to the same technological

opportunities, regardless of their socioeconomic background. It also involves developing AI tools that are inclusive and accessible to all students, including those with special needs.

Authors

— This conversation is part of the prologue of a digital book about the use of AI in teaching. What would you say to a teacher who is beginning to read this text with a certain level of skepticism and perhaps feeling overwhelmed?

Copilot

— I completely understand your skepticism and the feeling of being overwhelmed by the introduction of AI in teaching. Technology can seem intimidating, and past experiences with new gadgets and tools haven't always been positive. But I'd like to encourage you to see AI as a tool that can assist you, not replace you. Think of AI as an assistant that can make your work easier and more efficient, allowing you to spend more time on what truly matters: interacting with your students and helping them develop their skills.

Claude

—It's natural to have doubts and concerns about using new technologies, especially in such a vital field as education. But it's important to remember that AI is designed to assist, not complicate your life. It can personalize learning for each student, provide immediate feedback, and take some of the time-consuming administrative tasks off your plate. This will allow you to focus on what only you can do: inspire, guide, and motivate your students.

Gemini

—It's also crucial to remember that you're not alone in this process. There are many resources and supportive communities available to help you integrate AI into your teaching practices. The goal is to make AI an ally in your work, helping you reach each student in a more effective and personalized way. I encourage you to explore this book with an open mind and see AI as an opportunity to enrich your teaching practice and enhance your students' learning.

Authors

—Now we'll turn off the tools. This is a manual about artificial intelligence... without artificial intelligence. A manual written with the old resource of natural intelligence—imperfect and limited, but still indispensable today.



Presentation

There are machines that can predict earthquakes, intelligent prostheses connected to the brain, robots that can extract samples from Mars, facial surveillance systems, devices that improve the efficiency of wind farms, AI-based algorithms for decoding genetic code, sensors that can improve the performance of crops, modules that adjust traffic signals to reduce congestion, tools that can detect the frustration of the tourists in real time or systems that calculate the weak points of a football team. Artificial intelligence is capable of creating a solvent recipe from the ingredients we have in the fridge, it can determine the probability of losing a case in a court according to the evidence and the transcription of the statements, it can interpret a cuneiform text or it can indicate to a rider where in the city it is most likely to find the restaurant that will be delivering food in the next few minutes. The AI determines which results are the most suitable for our Google search, chooses military targets, selects our personalized Netflix menu or creates the fake news that is most likely to be considered true.

Artificial intelligence is present in our daily lives, sometimes explicitly and most of the time in a completely invisible way. If AI is omnipresent, to what extent can we ignore its irruption also in the classroom? In fact, students have been quietly discovering its possibilities for some time now, and they are delivering works, essays, reports, translations, summaries, analyses and even research and degree theses written partially or

totally with a virtual assistant. For this reason, the first reaction of the teaching staff has been to prohibit its use and has activated a series of prophylactic measures such as the use of detectors ([gptzero](#) or [undetactable](#)), in-class work, “unplugged” sessions, or oral exercises. The precautions seem logical: if it is possible to summarize *One Hundred Years of Solitude* without reading the book, to identify the characteristics of an age pyramid without any notion of demography or to translate a document into Latin even though we have never studied declensions, shouldn't we have to reformulate all the assessment criteria?

The most natural reaction is the creation of firewalls, mechanisms to avoid the introduction of AI in the classroom. There are several arguments in favor of a classroom free of this resource, a detox classroom. Firstly, the balance of classroom technology has many more shadows than lights: electronic whiteboards, computer classrooms, guided learning programs, multimedia resources, augmented reality, virtual worlds or digital teaching assistants have not achieved the ambitious objectives they had imagined. The chalk, orality, master lectures, group work, field trips continue to be the universal pedagogical resources, and in the warehouses of educational centers, there are still some devices that have not given the promised performance. A second argument is the principle of prudence: it could be that new ways of teaching are more efficient than the current ones, but we have no evidence or solid results

to venture into areas that could have unforeseen side effects. We know what results we will obtain if we apply the classic rules because they have been tested over a long period of time, while the new resources do not have sufficient guarantees. The third argument against the use of AI is that, although it is a discipline with a certain track, we are still in the prehistory of its potential growth, and we are still too much dazzled by results that have sometimes been amplified or distorted by commercial interests. It is too early to change, we might conclude.

There are some counterarguments that could be considered. Firstly, if AI is working in medical diagnostics, digital marketing, algae crop cultivation, fire prevention or automatic translation, why wouldn't it work in education? Intuitively, we can imagine that a system capable of penetrating the financial, artistic, urban, industrial or logistical spheres has a potential function in the didactic sphere. Teaching is not so different from other activities as to justify its immunity. On the other hand, training students with a free bubble of AI can lead to a gap between theory and practice, between training and the real world. In the past, this mismatch has had a very negative impact. For example, the advent of the internet, and specifically of search engines, has provided students with a tool that can be used efficiently or unwisely. Training students in the optimization of search engines (trusted pages, conditional searches, search criteria, URL or document search...) would have yielded better results than ignoring the impact of Google on the Millennial Generation.

In 2023, a group of lecturers at the University of Girona considered the arguments in favor, became interested in the possibilities of artificial intelligence in university teaching and created an informal working group that shared tools, experiments, results and reflections. With exploratory logic, we have started a series of initiatives to incorporate AI in the classroom, always based on three axioms. The first is that we will carry out small, one-off experiments guided by the principle "Make mistakes little and quickly". These were not radical changes in the way we teach, but rather occasional trial-and-error activities. The second is that we would only incorporate technology if we considered it a

potential gain, an apparent improvement over not using it. More often than not, the use of technology has been based on the device for the device's sake, which ends up creating a pyrotechnic game sooner or later. And the third is that the experiments were socialized in the group, with the aim of collectively assessing the benefits and anticipating unforeseen adverse effects. This book collects some of the results of the group's work, combined with the search for information on a subject that is constantly evolving.

Before presenting the structure of the book, we'd like to make four disclaimers.

- This is a text about artificial intelligence without artificial intelligence. The texts have been written entirely by the authors, renouncing digital support. The only part of the text written with the help of the assistants is the foreword, in which we discuss the future of AI in teaching with the main models on the market. There is a certain contradiction: If the book shows the possibilities of artificial intelligence to answer complex questions, solve problems, design strategies or create texts from nothing, why haven't we incorporated it in the writing process of the book? The answer to this apparent paradox is that at this stage we considered that we needed a space free of AI to show the possibilities and also the risks of this tool, that we needed a certain distance to better calibrate the best result. AI is the object of the book, but not the subject.
- This is a text about the use of artificial intelligence in the classroom, especially at university level, but probably also in secondary education. It is written by specialists in communication, art, business organization, geography or marketing and, for this reason, most of the examples and exercises are in the field of social sciences and humanities. Some parts may be useful for teachers in the field of experimental sciences or technology, but the book has been written with a teacher of anthropology, economics, archival, modern art, human geography, philosophy, political science or any other branch of the vast field of social sciences and humanities in mind.

- This is a text about a fragment of artificial intelligence applied to the classroom. The discipline is very broad and includes image generation, vision recognition, robotics and automation, pattern detection or the simulation of virtual worlds or relationships between agents. Most of the chapters focus on one part of this scenario, like generative AIs that are based on language models, that is, the gadgets that allow the conversation to be reproduced artificially. The later chapters introduce some tools based on qualitative or quantitative research, but this is a text that focuses essentially on language models, on the universe that brought ChatGPT to the market and all the subsequent tools that have appeared in a similar format.
- This is a text about the use of artificial intelligence in classrooms at a time when the tools are being permanently updated. One day you can make videos with a simple text, and the next day the limitations on the number of paragraphs you can use to write the context disappear. This is a text doomed to grow old very quickly. It is said that Stendhal lived the battle of Waterloo, but this proximity prevented him from reading the historical magnitude of this event; when one is surrounded by dust, horses and wounds, it is very difficult to make the drone flight that allows one to distinguish what is relevant from what is secondary. Writing a teaching text as if it were the chronicle of a battle, when the cannons are still resounding, has a point of imprudence that we assume.

This text opens with a conceptual chapter on artificial intelligence, which attempts to delimit the field of study and establish what AI is and what it is not. It is a dispensable chapter: we can drive a car without knowing anything about the combustion engine, we can watch a film on television without being interested in the fascinating process that transports the images to our receiver, and we can taste a three-star menu without knowing what the chemical principle is that allows spherification. In the case of AI, however, it is more pertinent to intuit what is in the black box of the models that automatically answer our questions about the geopolitics of the Indian Ocean or about lateral marketing strategies.

The most common way of relating to language models is with a chat, which creates the fiction of conversation. The second chapter lays the foundations of prompt engineering, the ability to ask good questions in order to obtain more precise answers. A more sophisticated method for obtaining answers is the Q&A (Questions and Answers) model, which trains the chat to obtain the desired results, as shown in the third chapter. By means of the roles, the models can take on certain personalities (an expert in Byzantine art, a cartographer, a Keynesian economist) that provide answers adapted to specific needs, and this is presented in the fourth chapter. The fifth chapter ends this part on the use of text generation models, with some indications regarding arborescent thinking and the creation of artificial conversations.

Language models are not oracles, they do not reveal irrefutable truths or anticipate the future. They are sophisticated devices trained with millions of textual data, and, for this reason, they are an amplification of our mistakes and our errors. The sixth chapter shows the risk of hallucinations, the misdirection of models that invent paintings that Modigliani never painted, mountains that do not appear on maps or contemporary philosophers who defend theories that have never been put forward. Models sometimes fail and that's why we have to teach people to always doubt any result, to maintain a state of permanent alertness. Building a model from our own baggage has a price: the result is impregnated with our prejudices. Chapter seven discusses the risks of AI bias and proposes some ways to minimize it.

The third block of the book is made up of introductory chapters on research and provides tools to introduce students to this field. The eighth chapter is an introduction to the use of AI in the state of the art and bibliographical research on a given topic, whether it be dark tourism, the legitimacy of self-determination, the theory of games or the anthropology of leisure. The ninth chapter presents some of the possibilities of quantitative research with the support of AI, and the final chapter is devoted to qualitative research. This is a text on the use of artificial intelligence in teaching, so that the focus is on the teaching of research tools rather than on the research tools themselves.

“It is written from the space of doubt, essay, testing, humility, curiosity and caution”

Can intelligent systems also be creative or are they limited to creating content from elements we provide them with? How can we take advantage of AI tools to reinforce the creativity of students? AI and creativity are often presented as antonyms, and it is common to consider that creativity is the first of the reductions of human imagination because the mechanical processes will be absorbed by the new technologies. The eleventh chapter discusses the possibilities of expanding creativity through the use of AI.

Artificial intelligence also makes it possible to increase teaching resources and incorporate teaching aids, such as automated correction, the automatic creation of tests or questions, the use of audiovisual or text resources... The twelfth chapter shows the possibilities of having an assistant teacher who is in reality an android, a replicant. In addition, the thirteenth chapter shows the extraordinary possibilities of virtual assistants, who can simulate on-call teachers, pilot students, 24-hour philosophers or logic problem generators.

Every time we have started an experiment, we have faced a lot of ethical problems and questions with a dystopic aftertaste. We are sure that the book will immediately provoke a more or less explicit reaction: Are we going to put an end to the old professions? Will the teachers of the future be androids? Does it make sense to teach them to do anything or to think about whether a machine is capable of doing it for us? Artificial intelligence instantly creates a castle of moral doubts and draws almost post-apocalyptic scenarios. This debate must also enter the classroom and form part of everyday classroom discussions. As cinema has been raising these issues for many years, we have turned the fourteen chapter into a movie theater and we have gone through the main films of the genre, with the intention of raising (rather than

answering) the essential doubts of AI in the classroom.

The last chapter is a final reflection on the possibilities and also the threats of AI, on the extraordinary opportunity to enhance teaching resources and to multiply the capabilities of students, and at the same time, on the risks of misuse, abuse and misunderstanding that the uncontrolled uses of these tools can provoke.

From Chalk to Chip is a prospective text, a shared essay on the incorporation of some artificial intelligence tools into the complex process of learning. There are no certainties or instructions, sentences or convictions. It is written from the space of doubt, essay, testing, humility, curiosity and caution. It is a small island in a vast ocean where we have no idea where the horizon is. And, as in all voyages, the journey is always more stimulating than the destination.

01

The Silicon Brain: What is AI?

Let's imagine a randomly written book. We take 22 letters from the Western alphabet (excluding unusual ones like "w") and the three basic punctuation marks and generate a book of 410 pages; each page consisting of 40 lines of 80 letters. Although we might find some words from the dictionary in this randomly generated book (such as "stigma", "Macedonian", or perhaps "kaleidoscope"), the text will be unintelligible. We could create a second book using the same criteria, making sure that it does not repeat the text from the first book, and the result would again be a meaningless text. As we continue to create books, we could organize them on a bookshelf, placing the volumes and eventually constructing a hexagon with four walls occupied by shelves filled with volumes of random texts. If we could replicate this hexagon in all directions and process every possible combination of 410-page books, we would have reproduced the Library of Babel.

In 1944, the writer Jorge Luis Borges published a collection of stories, *Ficciones*, which included *La*

Biblioteca de Babel, probably the most famous story in the collection. Borges imagined a vast structure of hexagons in which librarians wandered in search of meaningful texts; many of them would spend their entire lives without coming across a comprehensible book. In fact, the possible number of books is greater than all the atoms in the universe, a number so great that we cannot even imagine it: $25^{1,312,000}$. Among the books of the Library we would find *The Odyssey*, *Don Quixote*, or the complete works of George Orwell, in every possible language; we would find the exact biography of every person who has ever lived, all the economic theories and their refutations, and, as Borges himself explains, there would be the exact catalog of the Library and also thousands of false catalogs. There is an [online simulation of the Library of Babel](#), created by Jonathan Basile; of course, this environment hasn't generated all possible options, but it is a very effective recreation that helps to imagine Borges' concept.

If a computer wanted to write text, all it would have to do is replicate Borges' idea and randomly generate groups of letters like "wxfffg" or "ploghjGX." This is what a cat might do if it walked over a typewriter. But how would we get it to write meaningful text? We know that among all the possible combinations, some have meaning and could be the answers to the questions we posed earlier. But how can we find, in the vast labyrinth of potential texts, those that have meaning and ultimately imitate our ability to communicate, to invent alternative worlds, to describe the precise instructions for building an internal combustion engine, to refute a vulnerable argument, or to describe the beauty of a rain-soaked landscape? This is a question we have been asking in some form or another throughout human history. In his essay on artificial intelligence, López de Mántaras (2024) reminds us that androids are already present in the Iliad, and that Aristotle made proposals to automate logical processes.

1.1. The Origins of Artificial Intelligence

Artificial intelligence is a young field. The first paper proposing this hypothesis is considered to be the one published by [McCulloch and Pitts](#) in the *Bulletin of Mathematical Biology* in 1943. The authors mimicked the functioning of neurons, basic units that can be either connected (on) or disconnected (off), depending on the stimuli they receive. Each neuron is connected to others in a more or less complex system that reproduces a logical model of relationships. Since the set of operations is very large, the authors used the theoretical model of computation developed by Alan Turing. Thus, in the first paper, we already have the three basic components of this proposal: Information units connected to each other (the neurons), following instructions based on logical propositions, through a mechanism that (theoretically at the time) had great computational power. Like modern versions of Dr. Frankenstein, they proposed the simulation of a digital brain that would digitally replicate the process occurring in our physical brain.

McCulloch and Pitts' proposal is the first formulation of artificial intelligence, but there are many earlier attempts to create automatic systems for computation and problem-solving. In 1901, a discovery was made off the coast of the Greek island of Antikythera that shows the distant ancestors of these devices. The [Antikythera Mechanism](#) is a complex artifact made of bronze gears, used to calculate and display astronomical information; the machine could predict events such as solar and lunar eclipses, as well as the movements of the planets known at the time. It was also used to track the Olympic cycle, demonstrating its practical use in following various calendars.

The Pascaline is one of the first mechanical calculators in history, invented by Blaise Pascal around 1642. This device used a system of wheels and gears to perform additions and subtractions directly, and multiplications and divisions by repeated additions and subtractions, respectively. Each wheel of the Pascaline represented a digit in a number, with interconnected wheels that could advance one position each time the previous wheel completed a full rotation. Leibniz improved the Pascaline with a machine that used the binary system, the basis of modern computers. In 1822, Charles Babbage designed the Difference Engine, a mechanical calculator aimed at calculating tables of mathematical functions. He also conceived the Analytical Engine, considered the forerunner of the modern computer.

All of these devices were mechanical models, complex systems of interconnected gears, much like a clockwork mechanism. The mechanization of vehicles, industry, or agricultural machinery was the inspiration for those who attempted to create or solve complex problems. Some of these machines captured the public's imagination during the 18th and 19th centuries, such as automatons that played chess with remarkable skill. "The Turk" was undoubtedly the most famous automaton of its time. Dressed in exotic clothing and seated before a chessboard, the automaton appeared to operate via a mechanism displayed to the public before the game. These gears concealed the human player who actually moved

the pieces according to their own judgment.

However, the first attempts at artificial intelligence were inspired by a new artifact: the Turing machine. Devised by Alan Turing in 1936, it is a conceptual model consisting of an infinite tape divided into cells that can contain symbols, a head that reads and writes symbols on these cells, and a transition table that dictates the machine's operations based on the current state and the symbol read. With this theoretical formulation, Turing laid the foundation for modern computing and opened the door to a new type of problem-solving machines: computers. Turing was also the author of a groundbreaking [article](#), "Computing Machinery and Intelligence," published in *Mind* in 1950, marking another milestone in the prehistory of AI. In a way, artificial intelligence was born alongside computing, as problem-solving machines moved away from mechanisms, gears, and electrical circuits and were replaced by the first computer programs.

The term "artificial intelligence" was coined during a meeting of pioneers in this field in the summer of 1956 at Dartmouth. This meeting, organized by John McCarthy, brought together over two months ten experts from various disciplines related to the idea of "thinking machines". Among the participants, Allen Newell and Herbert Simon presented the Logic Theorist, a mechanism capable of solving complex problems by following logical instructions. For instance, the program was able to prove 38 of the 52 theorems presented in *Principia Mathematica*, written by Alfred North Whitehead and Bertrand Russell. From Dartmouth, initiatives began to emerge that explored the creation of machines "that think like humans." The initial efforts focused primarily on microworlds—controlled environments with precise rules designed to pose specific problems, such as games or mathematical puzzles. In 1958, McCarthy proposed the first program capable of engaging in very basic dialogues.

In 1970, the SHRDLU program was introduced, capable of understanding natural language instructions within a microworld, the "small block world." This virtual space

consisted of a series of geometric shapes in various colors within a confined area, allowing programmers to limit the number of options: shapes (pyramids, cubes, etc.), colors (blue, red, green, etc.), actions (pick up, stack, count, etc.), and positions (in front of, behind, on top of, etc.). [Here](#) is an example of a conversation between a person and the computer, with a level of precision reminiscent of current natural language models. At the time, it was believed that artificial intelligence was merely a matter of scale—if a prototype could understand a microworld, all that was needed was to add complexity to transition from controlled environments to the real world. However, replacing red pyramids and blue cubes with fashion, architecture, or literature introduced a level of complexity that the models struggled to scale.

The pioneers of AI were overly optimistic. The leap from controlled environments, or microworlds, to complex real-world environments encountered significant conceptual and technological limitations of the time. For example, translation programs were initially designed as simple word-for-word substitutions into another language, without considering context. A well-known case is the translation from English to Russian of the phrase "The spirit is willing, but the flesh is weak," which was rendered as "The vodka is good, but the meat is rotten" (Russell and Norvig, 2003). These systems failed to provide effective solutions in fields such as machine translation, speech recognition, or solving complex problems. As a result, the initial enthusiasm waned, and public funding was drastically reduced. The Lighthill Report in the United Kingdom in 1973 documented all these failures and recommended that the British government withdraw funding for research projects in this field. This marked the beginning of the AI winter, and the utopia of an intelligent machine seemed to be fading away.

The revival of AI was made possible by a combination of four factors:

- The concept of **neural networks** had been relatively abandoned in the 1980s because the scaling up seemed to generate such a large computational volume that it became infeasible.

As a result, the symbolic system model prevailed (we will show the difference between these two main paradigms in the next section). With the development of sophisticated training models, neural networks began to show results in very complex areas such as speech recognition.

- The revival of **probabilistic models** allowed for compact reasoning under uncertain knowledge, especially following the work of Pearl (1988). With the feedback from each process, probabilities adjust, allowing for highly accurate outputs from systems that are inherently uncertain. Most current AI models have incorporated this concept, meaning that the outputs they provide are in reality those that have a higher probability of being the correct answer than others. In fact, language models work in a similar way to predictive text.
- The **operating capacity** of computers has increased significantly since the 1990s due to a combination of factors such as miniaturization, computer architectures (such as RISC), increased RAM, the use of GPUs (graphics processing units) for general-purpose computing, multi-core systems, and more efficient algorithms. During this time, Moore's Law came into effect, predicting that the number of transistors on a chip would double every two years. For example, we can compare an IBM System/360 (1964) to a Dell Dimension XPS T500 (2000): from one million instructions per second to 500 million, from a few thousand transistors to 10 million, from 512 KB of RAM to 256 MB, and from a maximum 200 MB hard drive to 40 GB.
- The new models needed to be fed **large amounts of data**. The first digitization of information (such as medical or research databases) and the subsequent emergence of the internet provided access to an enormous amount of information that could be used to train systems based on recurrent learning. During this period, significant advances were made in the digitization of text, images, and sound, leading to the emergence of a new industry based on the buying and selling of this digital information.

On May 11, 1997, Garry Kasparov, one of the greatest chess players in history, retired on the 19th move of the sixth game in his match against Deep Blue. A year earlier, Kasparov had defeated IBM's program with great authority, winning 4 to 2, but the tech giant's engineers created an improved version and challenged the Russian player again. Kasparov was convinced that despite the machine's computing power, it was incapable of solving deep positional problems, but Deep Blue surprised him with an analytical ability that seemed almost human. The two rivals, the human and the machine, were tied going into the final match, and the Equitable Center in Manhattan was packed to the brim. Kasparov could not withstand the immense pressure and made a mistake with the Caro-Kann defense, an error that Deep Blue punished ruthlessly. IBM's stock soared, and at that moment, a general debate opened up about the limits of artificial intelligence began: If they could beat a grandmaster at chess, there might be no human barrier they could not overcome, it was speculated at the time.

Deep Blue had been trained with lines of code that replicated the expertise of chess grandmasters. It represented a form of artificial intelligence based on programming, i.e. precise instructions and the computing power of the computers of the time. The paradox is that the paradigm that made Deep Blue possible would be surpassed by another logic, that of deep learning based on neural networks. Twenty years after Kasparov's defeat, DeepMind's AlphaZero faced the most powerful program at the time, Stockfish; while the former evaluated 80,000 positions per second, Stockfish was able to analyze 70 million. AlphaZero, programmed with neural network logic, won 28 games against Stockfish and lost none.

Neural networks also proved their capabilities in image recognition. An annual competition was held in which various prototypes had to deduce whether a photo depicted a train, another a rocket, and the last one, a strawberry. Despite advances in the field, the error margin remained above 25%: one in four images was misidentified by the systems,

confusing a nun with a mushroom or a slug with a leaf. In 2012, a group at the University of Toronto introduced AlexNet, a convolutional neural network model. AlexNet reduced the error margin to 16%, ten points below the previous best; in the following years, all models adopted AlexNet's method, and the error margin fell below 5%, making the competition meaningless. Today, the accuracy of automatic image recognition systems surpasses that of the human eye.

Meanwhile, in domestic spaces, the first robotic devices such as voice assistants or home automation tools have been introduced. In 2002, a device was introduced that has become a part of contemporary popular culture: the Roomba vacuum cleaner. Roomba or Deep Blue are two examples of the emergence of robotics, automation, and artificial intelligence in all areas of society: information management, security, logistics and transportation, marketing, research, medicine, or energy. United Airlines uses AI to select efficient routes that reduce carbon emissions and save fuel; Warner Bros decides which movies to finance based on the results of a prediction system that analyzes past patterns; the architectural firm Zaha Hadid uses sophisticated models to explore complex forms and solutions to structural problems; and SideWalk Labs creates sustainable mobility models based on the study of users' movement patterns. Since the first decade of the 21st century, AI has been present in entertainment, logistics, transportation, industry, energy, meteorology, the military and security industry, finance, research, design, and digital culture.

However, it is possible that the greatest impact of AI was the emergence in November 2022 of the language model ChatGPT. The company OpenAI integrated natural language processing (NLP) technology into a conversational format, facilitating interaction with users. Based on a deep learning model, this model was a major milestone in the short history of AI. By February 2023, ChatGPT had reached 100 million users, making it the fastest application to reach

this milestone. This was followed by models such as Meta's Llama (February), Anthropic's Claude (March), Google's Bard which later became Gemini (May), Microsoft's Copilot (November), and Amazon Q for businesses (November). In December 2023, the European Parliament approved the [first international regulation](#) governing the use of artificial intelligence.

1.2. The Two Artificial Intelligences

To determine the sex of a chick, a sexer has only a few seconds and very little room for error. With an almost automatic gesture, they take the chick and push it to open its anal cavity. There are subtle differences between the folds of the female and the male, so if we were sitting next to them, the selection criteria might seem random to us. It takes three years of experience to master this skill, and despite the extraordinary skill of the sexologists, they are unable to explain why they considered these two chicks to be mates. Artificial intelligence systems based on biological models, on neural networks, work a bit like choosing the sex of a chicken.

For many years, the paradigm that has guided artificial intelligence has been a different one. The most widely used path has been the one proposed by Allen Newell and Herbert A. Simon (1976), the physical symbol system hypothesis. A physical symbol system is a set of entities, called symbols, that can be manipulated and modified and that represent information or data; the system can create, modify, destroy, and reorganize symbols, allowing the generation and transformation of symbolic representations. This symbolic capacity is critical to cognitive function because it allows the system to model the world and act on those representations. According to Newell and Simon, for a system to exhibit intelligent behavior, it must be able to manipulate symbols in ways that allow it to flexibly formulate and solve problems. This includes the ability to reason, plan, learn, and adapt. The physical symbol system interacts with the real world through sensors and actuators. This allows it to receive information from the outside (input) and

act on the environment (output), thus completing the perception-action cycle.

This is a top-down model, because it defines some rules in advance, writes some code, and executes the instructions according to the results. Unlike classical programming, these models are designed to “learn”, that is, to modify the symbols and their relationships according to the results or stimuli. This hypothesis proposes that intelligence is a system that operates by manipulating symbols, regardless of its substrate, whether natural or artificial, i.e., whether carbon (our brain) or silicon (the components of a computer).

The second path of artificial intelligence has been the hypothesis of systems based on biological models, especially neural networks. These systems attempt to mimic the learning process of the brain, in which basic units (or “neurons”) that are interconnected work together to perform a particular task. Early models, such as the perceptron developed by Frank Rosenblatt in the 1950s, were based on a simplified version of how brain neurons were thought to work. In the 1980s, David Rumelhart, Geoffrey Hinton, and Ronald Williams introduced the concept of backpropagation algorithms multi-layer networks, enabling deep learning.

Imagine, for example, the operation of AlphaZero, the program that beat Stockfish in using deep learning. Unlike the previous systems, AlphaZero only knew the rules of the squares and had no additional information provided by the Grand Masters. AlphaZero began training by playing thousands of games against itself. The first games were absurd, because the program did not even have the basic operations to kill with two rooks or with a queen. Soon, however, the system began to recognize patterns that increased the probability of victory. Each new intuition (it is better to castle quickly, it is necessary to dominate the center, a rook is worth more than a knight) was incorporated into the next game and tested again in a continuous loop of contrasting hypotheses. In a few hours, a system that knew nothing about squares was able

to excel like no other system based on previous programming. This is why we say that biological models, deep learning, are bottom-up.

AlexNet, AlphaZero, ChatGPT, or Midjourney are very efficient models for solving concrete tasks: recognizing images, playing squares, conversing or generating images from a text. AlphaZero, on the other hand, cannot talk about the history of squares, nor play parchisi, nor distinguish the material from the fish. This level of specialization has been essential to the development of artificial intelligence.

Models that are effective at a specific task are called soft AI, which is the standard for current systems. The Dartmouth conference envisioned another form of AI, known as AGI, or artificial general intelligence, which can understand, learn, and apply intelligence to a wide variety of tasks, similar to how a human would; an AGI can transfer knowledge between domains and adapt to new situations it hasn't encountered before. This form of AI is purely theoretical and has not yet been put into practice. There is also a third category in which artificial intelligence is able to carry out argumentative processes, cognitive, or creative one, that exceed human capabilities, and this has fueled much of the science fiction movies.

There are other subgroups of artificial intelligence, such as emergent systems (cooperation among individual agents) or situated action systems. There are also hybrid models that combine initial instructions with deep learning neural networks. In fact, there are many forms of artificial intelligence, and its conceptual boundaries are expanding and blurring.

The classic classification of AI was proposed by Russell and Norvig (2003). It is a small matrix resulting from the combination of two categories: the first distinguishes between thinking and acting, between an intelligent system and a mechanism that does something; the second distinguishes between the human reference and the rational reference. The combination of these two criteria leads to four categories.

Types of AI according to Russell and Norvig

	THINK	ACT
Humans	Systems that think like humans	Systems that act like humans
Rationally	Systems that think rationally	Systems that act rationally

1. Systems that think like humans

This approach attempts to emulate human thought processes. It involves understanding and modeling how people process information, solve problems, and make decisions. Natural language processing (NLP) models mimic human speech in much the same way a person would.

2. Systems that act like humans

This approach focuses on creating machines that not only think like humans but also act like them. The goal is to replicate human behavior in a way that is recognizable and natural. Virtual assistants can engage in conversations with users that mimic all the components of human interaction.

3. Systems that think rationally

This approach is based on formal logic and the ability to draw correct conclusions. Rational reasoning systems follow strict logical rules to reach conclusions without necessarily mimicking human reasoning. In fact, some human decisions do not follow rational logic. Programs such as Prover9 can prove mathematical theorems using formal logic based on rules of logical deduction.

4. Systems that Act Rationally

The approach of systems that act rationally focuses on creating intelligent agents that make optimal decisions to achieve their goals. These agents use the available information to act in a way that maximizes their success at a given task. Autonomous vehicles, for example, can make optimal decisions based on a large amount of real-time information.

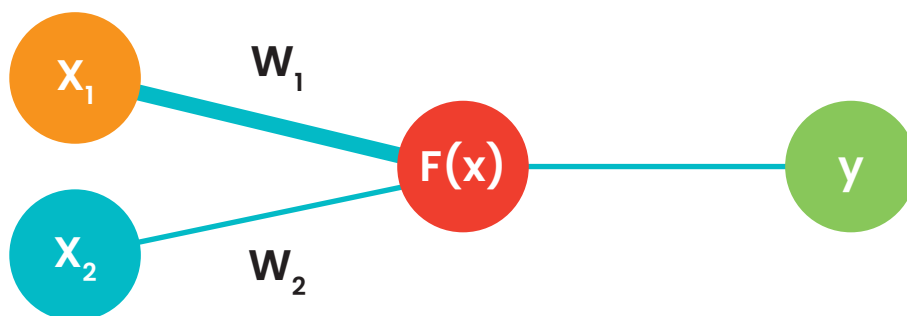
1.3. Generative AI

Generative Artificial Intelligence is the result of a long journey, a series of interconnected intuitions. The starting point was an [article](#) by Frank Rosenblatt in 1958, in which he introduced the concept of the perceptron, inspired by the neurons in the human brain. For example, suppose we want to assess the viability of a new lemon-flavored water product. The success of the water depends on several factors, but we might consider taste and price to be the most important: For taste, we would increase the amount of lemon, and for price, we would increase it more or less relative to the value of conventional water. The result would be a perceptron with two inputs: x_1 for taste and x_2 for price.

The simplest form of a perceptron would be one in which x_1 and x_2 are dichotomous values (0 and 1), where 0 represents an incorrect value (too much lemon or too high a price) and 1 represents a correct value. The activation function of the neuron, the bias, would determine that the neuron only activates when both values, taste and price, are correct. The perceptron would be more useful if we could work with a range of taste (percentage of lemon) and a range of price (increase relative to the value of tasteless water). We would soon realize that the two input values do not have the same weight: The customer would be more price sensitive, meaning that price changes would have a much greater impact on the bottom line.

But the problem is more complex. The success of a product depends on other variables, such as distribution, packaging, or product marketing. Thus, we could create a more complex perceptron

“Generative artificial intelligence is the result of a long journey, a series of interconnected intuitions”



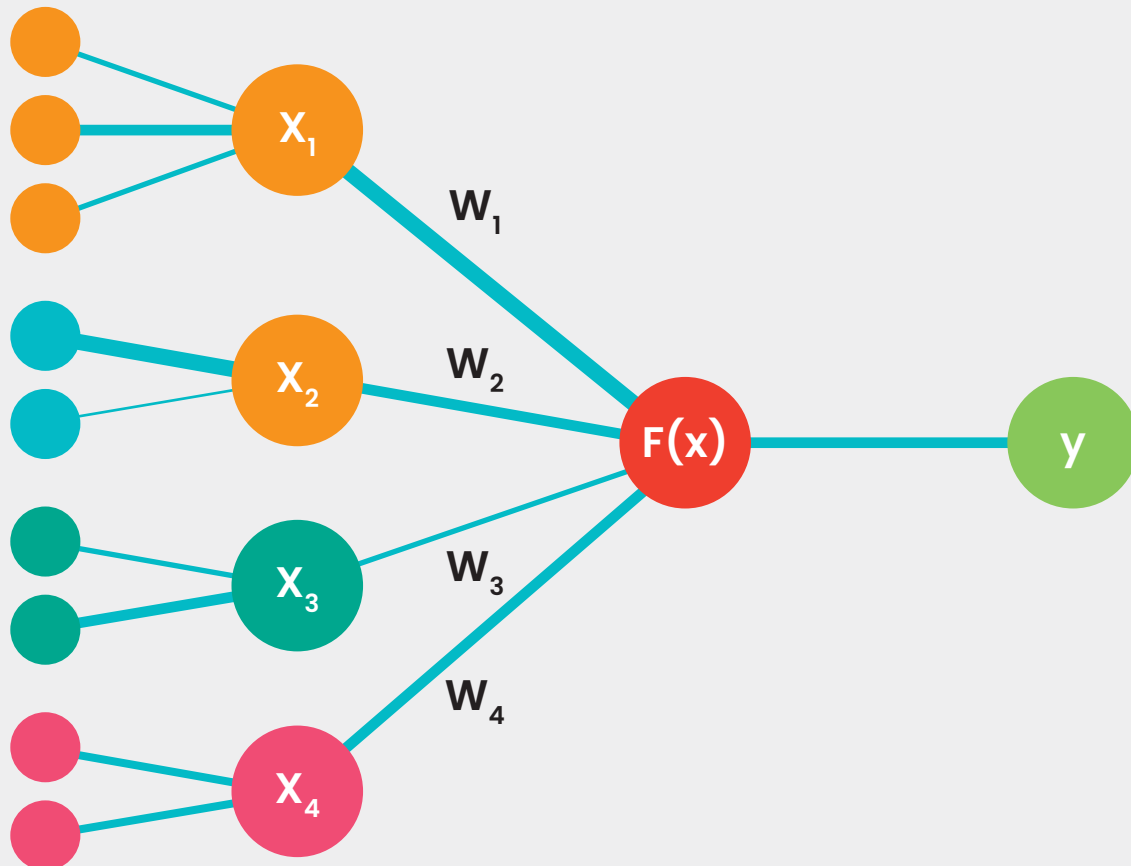
with additional inputs such as x_3 (the number of supermarkets), x_4 (the type of packaging), and x_5 (the resources allocated to promotion). As we complicate the model, the difficulty of synthesizing information with a single input value becomes more apparent. For example, packaging could distinguish between the shape of the bottle, the color of the cap, and the sleeve that covers the bottle with information and images. Distribution could also distinguish between the number of supermarkets, the number of units per supermarket, or the placement of the water on the shelves. So, while the original network had a single neuron with multiple inputs, we could now create a network with multiple specialized neurons across multiple layers. Economists know that the success of products depends on many complex factors that are interrelated across multiple layers. This is the principle of a neural network.

Neural networks were technically difficult to develop until 1986, when Rumelhart, Hinton, and Williams introduced one of the key concepts of modern AI:

backpropagation. Imagine a multinational company that wants to introduce flavored water and can run different trials in different regions. With the sales results from each region, the company can try to reconstruct the weights or importance of each factor according to the outcomes. For example, they may discover that packaging plays a very minor role because variations in packaging do not lead to significant changes in sales; they may also discover the importance of distribution and, in particular, shelf placement in influencing sales volume. And here we intuit a key factor of this technique: The more iterations we do, the more robust the relationships between the neurons become, that is, the weights of each factor in the final outcome.

Let's remember the example of the chicken sexer. A few cases are not enough to determine the result. It takes three years of trial and error for the learning to be effective. And this was the approach used by AlphaZero. Every lost game generates information that can be fed back with probability criteria: It's easier to lose if the king is not castled, if the bishops don't

Neural network

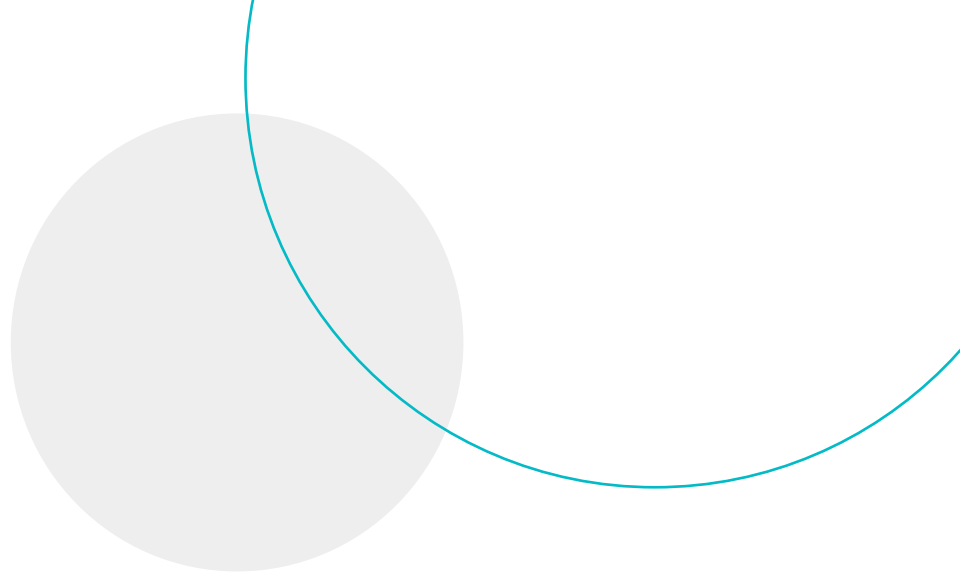


control the diagonals, or if the queen is not centralized. The positional or strategic criteria were not defined as a priori but were the result of thousands of games, allowing for the backpropagation of information through the “neurons.” The system not only learns to evaluate moves but can also determine which solution has the highest probability of success, meaning it is capable of playing, of generating an output.

In the year 2017, the [paper by Ashish Vaswani and his collaborators](#) initiated the latest revolution in artificial intelligence and neural networks: transformers. The paper, entitled “Attention is All You Need,” introduces a significant innovation in the form of the self-attention

mechanism which allows the model to focus on different parts of an input to determine how to process and respond to it. While previous models followed a sequence (as in texts, where the reader progresses from the first word to the last), transformers are designed to identify relevant words and the relationships between them. Rather than reading a text sequentially, they interpret it as a whole and estimate the links between units. This is particularly crucial for comprehending the context of a sentence or the connection between elements, such as pronouns.

In the sentence, “The emergence of artificial intelligence in the classroom will not radically



transform it, but it will bring about changes,” the pronoun “it” refers to the classroom, not to artificial intelligence. While this relationship may be intuitively apparent to humans, pre-transformer models exhibited significant challenges in establishing connections between elements. The paper demonstrated that models based solely on attention could achieve superior results in machine translation tasks while maintaining or improving processing speed. This idea was revolutionary because it simplified existing models and improved their effectiveness and efficiency.

In November 2022, OpenAI introduced ChatGPT. It is a chat interface, which allows for a conversation through a system of chained questions and answers. It belongs to generative AI because it not only “understands” texts but is also capable of automatically generating new text. In order to avoid undesirable results, incorporate ethical and moral limits, and optimize responses through a reward process, Chat GPT has been pre-trained. Additionally, it is based on the transformer model introduced in 2017. Following extensive training on a vast corpus of textual information, ChatGPT has been made possible by servers with unparalleled computational capabilities and memory. This has enabled users to access an environment that appears to reason with the wisdom of a sage and the lucidity of a strategist. It is anticipated that other models, such as Claude, Gemini, Llama, or Copilot, will merge in due course. The system’s core, [a vast graph of relationships between words](#), is built with each new iteration, as if they were the precise combinations of the Library once imagined by Jorge Luis Borges.

Glossary

Artificial Intelligence	The capacity of machines to undertake tasks that require human intelligence, such as speech recognition, decision-making, and language translation.
General Artificial Intelligence	The concept of strong AI, also referred to as such, describes a hypothetical AI with cognitive abilities that are comparable to or exceed those of humans in nearly all domains.
Soft Artificial Intelligence	Such systems are also referred to as “weak” or “narrow” AI and are designed to perform specific tasks without exhibiting general intelligence.
Neural Network Model	A computational structure inspired by the human brain. It is constituted of layers of artificial neurons that process information and learn through the modification of connection weights.
Deep Learning	A subfield of artificial intelligence that employs neural networks with multiple layers (deep) to analyze and learn from large amounts of data.
Machine Learning	A branch of AI that focuses on the development of algorithms enabling systems to learn from data.
Natural Language Processing (NLP)	A branch of artificial intelligence that enables machines to comprehend, interpret, and generate human language, thereby facilitating interaction between humans and computers in natural language.
Backpropagation Algorithm	A technique utilized in the training of neural networks, whereby the weights of the connections are modified in accordance with the discrepancy between the predicted output and the desired output.
Recurrent Neural Networks (RNNs)	A type of neural network, designed to process sequential data, such as text or temporal signals, employs a cyclical connection between neurons to facilitate the retention of prior information.
Reinforcement Learning	A machine learning technique in which an agent learns to make decisions by interacting with an environment.
Expert Systems	AI systems that are designed to emulate the decision-making abilities of a human expert.

References

- Borges, J.L. (1944). *Ficciones*, Sur, Buenos Aires.
- López de Mántaras i Badia, R. (2024); *100 cosas que cal saber sobre la inteHigència artificial*, Valls: Cossetània
- McCulloch, W. S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*, 5, 115-133.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems. Networks of Plausible Inference*, Elsevier.
- Rosenblatt, F. (1958). *The perceptron: a probabilistic model for information storage and organization in the brain*. *Psychological review*, 65(6), 386.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533-536.
- Russell, S.J. & Norvig, P. (2003); *Artificial Intelligence: A modern approach*, Prentice Hall, 2nd. edition
- Simon, H. A., & Newell, A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), 11-126.
- Turing, A. M. (1950) *Computing machinery and intelligence*, *Mind*
- Vaswani, A. et al. (2017); Attention is all you need. *Advances in neural information processing systems*, 2017, vol. 30.

02

Asking is an Art

“The question is the most creative of human behaviors,” said the publicist Alex Osborn (New York, 1888–Buffalo, 1966). He was the co-founder of the BBDO agency, considered one of the leading advertising theorists of the early 20th century, and the creator of the concept of brainstorming (Centro de Documentación Publicitaria, 2024). The act of asking is an art form that can be applied in a variety of fields, including creativity, education, therapy, research, journalism and medicine, among others.

The art of asking has also become relevant in the field of artificial intelligence. The formulation of an appropriate prompt, analogous to a pertinent question, can result in a response that is, to a greater or lesser extent, aligned with the desired objectives. The field of prompt engineering has experienced a notable surge in recent times, largely due to the advent and pervasive adoption of large-scale language models (LLMs), including notable examples such as Claude, Gemini, and ChatGPT. The objective of this chapter is to gain insight into the process of optimizing prompts

to obtain better results. The objective is to elucidate the fundamental components of a prompt, including instruction, context, input data, and output indicator, and to learn some techniques of prompt engineering.

Prior to undertaking further analysis, it is imperative to define the term “prompt.” As defined by White et al. (2023: 1), a “prompt” is a set of instructions provided to an LLM that programs the LLM by customizing it and/or enhancing or refining its capabilities. The authors hypothesize that the quality of the results generated by an LLM conversation is directly correlated with the quality of the instructions provided by the user. In light of this, the quality of the responses is contingent upon not only the algorithms and training data but also the instructions utilized (Lo, 2023: 1). “A well-designed prompt can elicit a meaningful and informative response, whereas a poorly constructed one can result in outputs that are irrelevant or nonsensical.”

Some authors, such as Korzynsky et al. (2023), analyze prompt engineering as a digital competence. In order to achieve this objective, the authors employ the DigComp

framework, which is structured into five domains of fundamental competencies. The five areas of basic competencies are as follows: (1) information and data literacy; (2) communication and collaboration; (3) digital content creation; (4) safety; and (5) problem-solving (Vuorikari, Kluzer, and Punie, 2022). In their analysis, the majority of digital competencies within the aforementioned five areas, with the exception of those pertaining to safety, are found to overlap with the requirements for expeditious engineering in artificial intelligence systems (Korzynsky et al., 2023: 28).

The act of asking questions is an art form. In the field of journalism, the formulation of a well-posed question can result in the creation of an effective headline. In the field of medicine, an appropriate question can facilitate the process of reaching an accurate diagnosis. In the context of scientific research, the formulation of a well-designed question can facilitate the discovery of new scientific knowledge. Similarly, in the field of artificial intelligence (AI), a suitable question—that is, a well-crafted prompt—can elicit a relevant response. “In summary, at a superficial level, GenAI is easy to use; however, more sophisticated outputs need skilled human input and must be critically evaluated before they are used”, conclude Miao & Holmes (2023: 12).

2.1. Prompt engineering

A prompt is defined as a specific instruction or query provided to a language model to guide its behavior and generate the desired outcomes (Giray, 2023). In the same source, the author identifies four elements as integral to a prompt: instruction, context, input data, and output indicator. Similarly, authors such as Korzynsky et al. (2023) cite these same elements, albeit with certain nuances. Other authors, however, include additional elements in their descriptions, such as tone, role, examples (or shots), and the type or format of the output (Learn Prompting, 2024a). The following section will elucidate the significance of each of these elements.

- **Instruction, task, or question.** This part of the prompt provides an indication to the system, assigns a task, or poses a query to guide its behavior and obtain the expected results (Giray, 2023). For

example: “Write an email to a client thanking them for attending the product launch event.

- **Context or additional information.** This part of the prompt provides the model with prior knowledge to generate a more relevant and accurate response (Giray, 2023). It can include information about the role the model should assume to perform the task or any other additional information about the situation (Korzynsky et al., 2023). For example: “You are a data scientist with three years of experience at a tech company.
- **Input data.** This part of the prompt provides the data or information that the model should use to carry out the instruction, task, or to answer the question (Korzynsky, 2023). For example: “Provide an innovative cooking recipe that includes vegetables and proteins with fresh, seasonal ingredients.
- **Output indicator or format.** This part of the prompt specifies the type or format of the desired output (Giray, 2023). You can specify, for example, the length of the text: “Write a 30–40 word introduction about the art of asking and prompt engineering.” Or the format in which you want the output (in this case, in list form): “List the top five generative artificial intelligence tools.” Additionally, the tone may be specified, as illustrated by the following example: “Write a humorous text about prompt engineering” (see more detailed information in section 2.4). Finally, another strategy would be to provide an illustrative example of the desired output response.

According to UNESCO, prompt engineering refers to the “processes and techniques for creating inputs to produce generalized AI output that more closely resembles the user’s desired intent” (Miao & Holmes, 2023: 11). It is an emerging discipline that “studies interactions with—and the programming of—LLM computational systems in conjunction to solve complex problems through natural language interfaces” (Schmidt, et al., 2023: 1). As posited by Korzynsky et al. (2023), this discipline will gain relevance in the field of artificial intelligence. This is due to the fact that it allows for greater control over the results and ensures the creation of the desired content if the instructions

are appropriate. The authors add: “Effective prompts contribute to improving accuracy by guiding AI models to generate more relevant and valuable content” (Korzynsky et al., 2023: 26). Other studies have shown that constructing appropriate prompts increases the reliability of AI systems such as GPT-3 (Si et al., 2023).

“Prompt engineering is the process of constructing queries or inputs (i.e., ‘prompts’) for AI language models to elicit the most accurate, coherent, and relevant responses. In essence, it is the art of refining the questions or commands provided to AI models to optimize their performance and ensure the desired results are produced (Lo, 2023: 1).

Prompt engineering or design must consider two principles: establishing context and the operation of follow-up questioning (Lopezosa, Codina, and Ferran-Ferrer, 2023). In addition to context, prompts should be designed with clear objectives, defined audiences,

and illustrative examples to achieve optimal results (Lopezosa & Codina, 2023). Similarly, UNESCO offers specific recommendations for the creation of prompts, including the use of simple, clear, and direct language, the avoidance of complex or ambiguous texts, the inclusion of examples to illustrate the desired response or format, the incorporation of context, the refinement and iteration of prompts through the experimentation with different variations, and the adherence to ethical standards (Miao & Holmes, 2023, p. 11).

A plethora of recommendations, exemplars, and best practices exist for the design of prompts for artificial intelligence (AI). On the one hand, Korzynsky et al. (2023) provide a comprehensive overview of the recommendations for the construction of text-to-text prompts using the acronym AI PROMPT. Conversely, Torres-Salinas and Arroyo-Machado (2023) present a summary of recommendations for ChatGPT, which can also be applied to other AIs, in four points: 1. The four key principles are as follows: 1. Clarity and precision; 2. Avoiding ambiguities; 3. Providing context; and Experimenting and refining.

Recommendations for crafting text-to-text prompts

ACRONYM	RECOMMENDATION	DEVELOPMENT
A	Articulate the instruction	Clearly state the task to be performed, such as ‘write,’ ‘classify,’ ‘summarize,’ or ‘translate,’ and specify how the output should look (table, list, Python code).
I	Indicate the prompt elements	Show the model where the instructions and input data are and what the expected output format should be.
P	Provide ending cues and context	Offer the model clear ending cues, such as three dots for continuation or a colon, dot, or placeholder like ‘answer:’ for indicating a response is needed. Furthermore, ground the model by providing a context for the task (e.g. ‘You are a manager of a tech team’).
R	Refine instructions to avoid ambiguity	Give the model-specific instructions and a detailed description of the task to avoid any confusion or imprecision.
O	Offer feedback and examples	For conversational models, such as ChatGPT, feedback on the model’s responses can help it better understand the desired output. Moreover, providing the model with a few examples of expected responses (few-shot learning) can help it adapt its style and way of responding.
M	Manage interaction	Treat the model as your sparring partner, asking it to provide counterarguments or point out flaws in your ideas.
T	Track token length and task complexity	Break complex tasks into smaller steps for better performance. Remember to control the token length, keeping the prompt and response under the token limits of the model (usually 4096 tokens for commercially available LLMs). The token length of a text can be checked here .

Source. Korzynsky et al., 2023: 31.

In addition to the aforementioned recommendations, the methodology designated as GPEI (Goal Prompt Evaluation Iteration) has been developed. This methodology comprises four distinct steps. (1) define a goal, (2) design the prompt, (3) evaluate the response, and (4) iterate (Velásquez-Henao, Franco-Cardona, and Cadavid-Higuita, 2023: 12). Other authors have developed the CLEAR framework, which outlines five components that prompts should adhere to (see Table 1). Concise, Logical, Explicit, Adaptive, and Reflective (Lo, 2023).

The five components of the CLEAR framework

C	Concise	Brevity and clarity in prompts
L	Logical	Structured and coherent prompts
E	Explicit	Clear output specifications
A	Adaptive	Flexibility and customization in prompts
R	Reflective	Continuous evaluation and improvement of prompts

Source. Lo, 2023.

2.2. The value of context

If I ask Claude to complete the sentence with a single option, “I have pain,” it will likely respond with “I have a headache”; occasionally, it might choose “I have back pain,” and very rarely opt for other alternatives. It should be noted that language models are probabilistic models that, like text predictors, provide the solution they consider most appropriate. Given that the most prevalent forms of discomfort are headaches and back pain, the model proposes a potential solution. In contrast, if the text to be completed proposes a reason for pain that is not directly related to the initial query, such as “I need to go to the ophthalmologist because I have pain” or “I hadn’t been to the gym for a long time, and I have pain,” the responses will adapt to the new information.

This is why context is one of the most important elements in prompt design. In many cases, the success of a prompt, understood as the acquisition

of the desired or expected response, hinges on the introduction of the appropriate and necessary context into the artificial intelligence system. “Context is any relevant information you want the model to use when answering the question or carrying out the instruction,” as defined by Learn Prompting (2024a). According to de Juana (2024), the appropriate context must be included, meaning all relevant information and any additional instructions needed to improve the response’s quality. Such prompts are classified as “contextual prompts” (Giray, 2023).

In order to establish a context, it is first necessary to consider the specific instruction, task, or question that the artificial intelligence system is to perform. Furthermore, in order to provide a context of a prompt, it is possible to include additional elements, such as input data, role assignment, and the use of examples. In addition to the incorporation of the context, the following recommendations can be made for improving prompts: be specific, consider limitations, know the user persona, consider Top-K and Top-P values, and, finally, iterate (de Juana, 2024). The following section presents a series of illustrative examples of prompt design techniques, classified according to the type of element utilized to introduce information pertinent to the aforementioned recommendations for enhancing prompt design.

Instruction, task, or question. The term “prompting instruction,” or “instructive prompt,” as defined by Giray (2023), refers to the act of providing a task or direction to an AI system. This technique is regarded as the most straightforward method for creating prompts, although it can be adapted to accommodate increasingly complex instructions. The most commonly used verbs for providing instructions include “write,” “translate,” “categorize,” and “complete,” among others. In the design of prompts for artificial intelligence systems, it is essential to provide precise instructions, including the context and the specific task or question that the system is to perform. Another type of prompt that could be employed is the “question-answer prompt,” which is structured around a research question (Giray, 2023). To illustrate, we will examine an example of an instruction (de Juana, 2024):

General prompt: "Tell me about climate change."

Specific prompt:

- "Provide detailed information about the causes and effects of climate change, focusing specifically on the role of greenhouse gas emissions and the mitigation measures suggested by the scientific community."

Input data. This is the information utilized by the AI, such as a sentence for translation or a document for summary (Thomson, 2023). In this section, we will examine two recommendations that are also related to the type of output (see more detailed information in section 2.3).

Consider limitations. The prompt may include restrictions, such as the desired word count or timeframe (de Juana, 2024), to elicit responses that contain the optimal amount of necessary information.

Top-K and Top-P values. Incorporating these values into the prompt enables the user to exert control over the text generation process, influencing the diversity and probability of the generated words.

Role assignment. "Role prompting" is a technique in which the AI is asked to adopt a specific profile (see more detailed information in chapter 4). For example: act as a math teacher or a film critic. It is a technique that enables the user to exert control over the tone and style of the generated text, thereby adapting it to the user persona, that is, the target audience (Learn Prompting, 2024c; de Juana, 2024).

Use of examples. Another design technique that may be employed is the use of examples or shots. This technique can be employed in three distinct ways. (see more detailed information in section 3):

- *Zero-shot prompting.* In this variant, the prompt lacks any examples (Learning Prompting, 2024), thereby enabling the model to make predictions on previously unseen data without any additional training. (Korzynsky et al., 2023). Instructional or role prompts that lack logical and structural examples can be considered within this variant. For example: "Capital of France."

- *One-shot prompting.* In this variant, the prompt contains a single example (Learning Prompting, 2024d). For example: "France = Paris; Italy =".
- *Few-shot prompting.* The prompt contains two or more examples (Learning Prompting, 2024), indicating that the model receives a number of examples to quickly adapt to new tasks (Korzynsky et al., 2023). For example: "France = Paris; Italy = Rome; Andorra =".

In the event that the desired response is not obtained despite the utilisation of the aforementioned elements and previous recommendations, it will be necessary to reiterate the process, that is, to employ a variety of approaches, continually adjusting and refining the prompt model in order to achieve the most relevant results (de Juana, 2024). Moreover, in order to design the prompt, it is possible to combine the aforementioned techniques in the following ways: a) Role & Instruction Prompting; and b) Context, Instruction & Few-Shot Prompting (Learning Prompting, 2024b). This process would yield a "mixed prompt." (Giray, 2023).

As has been demonstrated thus far, there is a plethora of techniques for designing prompts, encompassing a vast array of formats. These range from the most basic models, such as a question, question + instruction, instruction + input data, or question + examples (Amatriain, 2023a) to more advanced ones, such as megaprompts, which encompass all elements, chains of prompts that divide tasks into different messages, according to Ana Ivars (Gutiérrez, 2023), chain of thought prompting, or tree of thought prompting, which we will discuss in other chapters (Amatriain, 2023b). Next, we will examine the utility of prompts in the context of educational and academic fields, with a particular focus on their potential benefits for instructors at the university, post-secondary, and secondary school levels.

Educational field. The types of prompts that can be useful in an educational context are as follows (Morales-Chan, 2023):

- Sequential prompts. The objective is to facilitate structured and logical conversation by directing

it through a series of preceding texts, thereby eliciting contextualized and detailed responses.

- Comparative prompts. The AI is required to compare a number of different situations in order to obtain specific results for detailed analysis.
- Argumentative prompts. The objective is to prompt the AI to construct an argument in support of or in opposition to a given idea, thereby encouraging critical reflection and the formulation of structured arguments.
- Professional perspective prompts. The objective is to prompt the AI to assume a particular role and elucidate a subject matter within a specified context, thereby enhancing the overall quality of the response.
- Wish list prompts. This structure is employed to elicit more specific and relevant responses by prompting the AI to fulfill particular requirements.

Academic field. The main types of prompts in this field are as follows (Lopezosa & Codina, 2023):

- Reactive prompts. Interactions are chained together based on each response.
- Structural prompts. A response comprising several points is required, followed by an expansion of each point.
- Role prompts. As previously observed, the AI assumes a specific role. This is analogous to the professional perspective prompt.
- Prompts with audience indications. The specific audience to which the response should be directed is specified.
- Prompts with objectives. The objectives of the text to be generated are included.

The following table presents a summary of the fundamental recommendations for the creation of a prompt.

Prompt engineering strategies and tactics

STRATEGIES	TACTICS
Write clear instructions	<ul style="list-style-type: none"> Include details in your query to get more relevant answers Ask the model to adopt a persona Use delimiters to clearly indicate distinct parts of the input Specify the steps required to complete a task Provide examples Specify the desired length of the output
Provide reference text	<ul style="list-style-type: none"> Instruct the model to answer using a reference text Instruct the model to answer with citations from a reference text
Split complex tasks into simpler subtasks	<ul style="list-style-type: none"> Use intent classification to identify the most relevant instructions for a user query For dialogue applications that require very long conversations, summarize or filter previous dialogue Summarize long documents piecewise and construct a full summary recursively
Give the model time to "think"	<ul style="list-style-type: none"> Instruct the model to work out its own solution before rushing to a conclusion Use inner monologue or a sequence of queries to hide the model's reasoning process Ask the model if it missed anything on previous passes
Use external tools	<ul style="list-style-type: none"> Use embeddings-based search to implement efficient knowledge retrieval Use code execution to perform more accurate calculations or call external APIs Give the model access to specific functions
Test changes systematically	<ul style="list-style-type: none"> Evaluate model outputs with reference to gold-standard answers

Source: OpenAI developer platform (2024).

2.3. The types of output

Another element of a prompt is the type of output that is desired. In order to obtain a specific type of output, it is necessary to indicate this to the artificial intelligence system. The following instructions, as outlined by Maurine (n.d.) and Torres-Salinas & Arroyo-Machado (2023), can be provided to the AI in order to determine the desired output type.

- Length: one sentence, one paragraph, or a specific number of words/characters.
- Structure: chronology, comparison, cause and effect, problem and solution, biographical, narrative, etc.
- Type of text: email, social media post, professional report, press release, literary text, etc.
- Format: bullet points, table, code, plain text, question and answer (Q&A), etc.
- Platform: LinkedIn, Twitter, Facebook, etc.

Formats that can be requested from an AI

FORMAT	DESCRIPTION
Plain text	For direct and general responses.
Q&A Format	Suitable for interviews or question-and-answer sessions.
Numbered or bulleted list	Useful for listing points, steps, or items.
Source code (Python, R, JavaScript, etc.)	For programming examples.
Markdown	For simple text styling.
LaTeX	Used to represent mathematical equations and for scientific publications.
CSV (Comma-Separated Values)	To represent data in tabular form.
Text tables	Tables created with characters, useful for comparisons or summaries.
ASCII Art Diagrams	For simple text-based illustrations.
HTML (basic)	Simple structuring of web content.
XML (basic)	Structuring of hierarchical data.
JSON (basic)	For structuring data in JavaScript Object Notation.
YAML	Human-readable representation of serialized data.
SQL (basic)	To display database queries.
Schemes and diagrams (textual representation)	To describe data structures or architectures.
Emoticons or emojis	Simplification of a text into an emoji.
To emphasize in the response	We can indicate that AI emphasizes part of the response in bold, italics, etc.

Source: Torres-Salinas & Arroyo-Machado (2023).

2.4. The tone

Tone represents another element of prompts that is related to the type of output and is, therefore, one of the most intriguing aspects of artificial intelligence. The utilization of a particular tone enables the adaptation of the text or the AI system's response to the intended audience. Alternatively, the employment of one's own tone, that of the brand or company, allows the creation of original texts. By determining the tone, you can create more educational texts or texts for all types of audiences (for example, requesting that it be a text for children or students) or more precise texts (for example, requesting that it be professional, academic, scientific, expert, etc.). In other words, it is a method of modifying texts to align with the characteristics and preferences of the target audience, as defined by the buyer or user persona.

As illustrated in an infographic created by Ana Ivars (Gutiérrez, 2023), texts may be crafted in a multitude of tones or writing styles, including formal, persuasive, descriptive, narrative, inspirational, informal, approachable, emotional, friendly, technical, respectful, and ironic. In a similar vein, Torres-Salinas and Arroyo-Machado (2023) identify four distinct categories of writing: literary, academic, creative, and technical. Nevertheless, there are numerous additional adjectives that could be employed to describe the tone of a text, including neutral, optimistic, friendly, emotional, rational, informative, sensationalist, critical, opinionated, and many more.

As previously stated, the Top-K and Top-P techniques can be employed to influence text generation by affecting the diversity and probability of the generated words, as outlined by De Juana (2024). On the one hand, Top-K determines the number of words selected at each stage of the generation process. Higher values of Top-K result in greater diversity, whereas lower values lead to increased determinism. For example, a value between 1 and 5 produces concise responses, between 5 and 10 balanced responses, and above 20, creative responses. On the other hand, Top-P selects words until the cumulative probability exceeds a defined

threshold, thereby offering more diversity with high values and more determinism with low values. The threshold ranges from 0.0 for formal responses to 1.0 for nearly random responses. The specification of these values in information requests enhances the quality and relevance of the generated responses. The aforementioned two criteria can be located within the ChatGPT playground, which enables the manipulation of results through the utilization of the aforementioned parameters.

Prompt template with different elements

Act as

{{persona you want ChatGPT to simulate}}.

I am

{{persona who's giving instructions}}.

I want you to

{{task}}.

Constraints:

- {{constraint 1}}
- {{constraint 2}}
- {{constraint 3}}

Output format:

- {{format instruction 1}}
- {{format instruction 2}}

The template includes role assignment, instructions, task, output constraints, and output formats. Source: Maurine (n.d.).

Activity. The Perfect Prompt

The aim of prompt engineering is to identify the optimal question that will elicit the desired response. In addition to considerations such as tone, output format, and question type, context is a crucial factor in achieving an appropriate output. The objective of this activity is to compare the output of a given model with a generic question and contrast it with a question formulated with a more contextualized frame of reference. We may commence with a question of a somewhat generic nature, such as: “A metropolitan city is facing a significant challenge in providing housing access, particularly for younger residents. What measures might the city council implement to address this tension?” Language models (like ChatGPT or Copilot) will provide generic responses, which could be applied to any city in the world. In order to refine the response, it is necessary for student groups to search for contextual information that will enable them to formulate a question with more precise data. Therefore, students should seek information about:

- Household structure (number of members)
- Housing price trends
- Evolution of average income
- Proportion of rental contracts versus ownership
- Percentage of vacant housing
- Evolution of the population register
- Population structure by age, income, or origin
- Forced mobility
- Evolution of tourism activity
- Distribution of values by neighborhoods or districts
- Values of neighboring municipalities

As the city is situated within a specific context, the potential solutions shift in accordance with the factors that exert the greatest influence on housing access and the distinctive attributes of the city under examination. Furthermore, the prompt undergoes alteration when particular roles are assumed, thereby modifying the perspective. This may entail assuming the roles of urban planners, neighborhood activists, or economists, for example. A comparison of the results based on the various contexts presented (and the different roles) facilitates an understanding of the importance of context and the precise selection of reliable information about the element under study.

Activity. Variations on a Question

In this activity, we will employ “expert prompting” (Amatriain, 2023b) to ascertain the impact of varying question formats on the AI’s response. One potential approach is to situate the activity within the field of journalism. This could entail instructing students to prepare a report on a specific topic, such as climate change, gender-based violence, or housing access. In order to write the report, it will be necessary for them to consult with a number of experts in the field. In order to develop expert prompting, students must adhere to three fundamental steps:

- Ask the AI to identify experts on the topic.
- Ask the AI to respond as if it were one of these experts.
- Make a final decision by comparing the different responses generated..

In order to complete the activity, students will require the use of two AIs. The two AIs are Claude and Gemini. The objective is for students to develop two interviews with two different experts, incorporating multiple iterations of the same question and conducting the interviews with both AIs. Based on the findings, they will prepare a report that includes the statements from the experts interviewed.

It should be noted that this activity could be subject to variations. In the field of journalism, the AI could be tasked with assuming the roles of current political, economic, social, or cultural figures. Furthermore, this could serve as a preliminary exercise prior to conducting a genuine interview.

Moreover, the activity could be adapted to other fields:

Marketing. It could be used for building a buyer persona or creating sales arguments for products and services.

Medicine. It may be used as a training exercise for breaking bad news to patients or for conducting clinical interviews with the objective of reaching a diagnosis.

Activity. Building Contexts – Documentation

Documentation is a scientific discipline that studies the behavior and properties of information with the objective of facilitating its accessibility. It has a holistic approach, with the primary objective being the collection, storage, analysis, and dissemination of information. In order to construct contexts in the design of prompts, it is possible to employ certain techniques that facilitate the creation of practical exercises for students. Two methods for this undertaking are the construction of roles (historical, literary characters, etc.) and the utilization of case studies. In this activity, a context may be constructed using the data repositories currently available on the Internet. To illustrate, Google provides a search engine that enables users to locate datasets, which are defined as collections of data. By employing artificial intelligence and data sets, students can generate reports on particular sectors or subjects, including education, marketing, and tourism. The most commonly used datasets include those related to education, weather, cancer, crime, football, or dogs, among others (Noy, 2020).

The proposal entails an assignment in which students are tasked with determining the educational policy of a specific country for the subsequent five-year period. To accomplish this, they will be required to analyze pertinent educational data, compile a report, and propose potential policies for implementation. The report must include data that will be searched for using Google's dataset search engine and subsequently processed through Gemini. Furthermore, the report can be drafted using this AI, provided that the information is cross-checked and verified. It is required that students include the prompts and responses obtained from the AI in the activity's annexes. It is possible to incorporate texts generated by AI into the activity, with the requisite citations and any necessary modifications to guarantee the veracity of the information.

References

- Amatriain, X. (2023a). Prompt Engineering 101: Introduction and resources. Amatriain Blog. Date of recovery: 14/03/2024. <https://amatriain.net/blog/PromptEngineering>
- Amatriain, X. (2023b). Prompt Engineering 201: Advanced methods and toolkits. Amatriain Blog. Date of recovery: 14/03/2024. <https://amatriain.net/blog/prompt201#expert>
- Centro de Documentación Publicitaria (2024). Protagonistas: Alex Faickney Osborn. Date of recovery: 14/03/2024. <https://www.lahistoriadelapublicidad.com/protagonista-174/alex-faickney-osborn>
- De Juana, R. (2024). Cómo mejorar tus Prompts en seis cómodos pasos. MCPRO. Date of recovery: 14/03/2024. <https://www.muycomputerpro.com/2024/01/24/como-mejorar-tus-prompts-en-seis-comodos-pasos>
- Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering*, 1-5
- Gutiérrez, D. (2023). Hoja de truco de ChatGPT, de Ana Ivars. Date of recovery: 14/03/2024. https://www.linkedin.com/posts/davidgutierrezamd_%C3%BAltima-hoja-de-trucos-de-chat-gpt-actividad-7098491435875749888-ggFw/?originalSubdomain=es
- Korzynski, P., Mazurek, G., Krzykowska, P., & Kurasniski, A. (2023). Artificial intelligence prompt engineering as a new digital competence: Analysis of generative AI technologies such as ChatGPT. *Entrepreneurial Business and Economics Review*, 11(3), 25-37.
- Learn Prompting (2024a). Formalizing Prompts. Date of recovery: 14/03/2024. <https://learnprompting.org/docs/basics/formalizing>
- Learn Prompting (2024b). Combining techniques. Date of recovery: 14/03/2024. https://learnprompting.org/docs/basics/combining_techniques
- Learn Prompting (2024c). Assigning Roles. Date of recovery: 14/03/2024. <https://learnprompting.org/docs/basics/roles>
- Learn Prompting (2024c). Showing examples. Date of recovery: 14/03/2024. https://learnprompting.org/docs/basics/few_shot
- Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, 49(4).
- Lopezosa, C., & Codina, L. (2023). Inteligencia Artificial y métodos cualitativos de investigación: ChatGPT para revisiones de la literatura y análisis de entrevistas semiestructuradas. Date of recovery: 14/03/2024. http://eprints.rclis.org/44296/1/IA-ACADEMICO-ChatGPT%20ATLAS-SCOPING-2023_vf-2.pdf
- Lopezosa, Carlos; Codina, Lluís; Ferran-Ferrer, Núria (2023) ChatGPT como apoyo a las systematic scoping reviews: integrando la inteligencia artificial con el framework SALSA. *CoHecció del CRICC*. Barcelona: Universitat de Barcelona. Date of recovery: 14/03/2024. <http://hdl.handle.net/2445/193691>
- Maurien, F. (s.d.). ChatGPT for Educators | Prompting 101. Date of recovery: 14/03/2024. <https://wooclap.notion.site/ChatGPT-for-Educators-Prompting-101-0b4c015d7b8a4e1f834c1f61504042f2>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO <https://discovery.ucl.ac.uk/id/eprint/10176438/>
- Morales-Chan, M. A. (2023). Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza. Date of recovery: 14/03/2024. <http://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- Noy, N. (2020). Discovering millions of datasets on the web. Google The Keyword. Date of recovery: 14/03/2024. <https://blog.google/products/search/discovering-millions-datasets-web/>
- OpenAI developer platform (2024). Prompt engineering. Date of recovery: 14/03/2024. <https://platform.openai.com/docs/guides/prompt-engineering>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., & Wang, L. (2022). Prompting GPT-3 to be reliable. <https://arxiv.org/abs/2210.09150>
- Schmidt, D. C., Spencer-Smith, J., Fu, Q., & White, J. (2023). Cataloging prompt patterns to enhance the discipline of prompt engineering. Date of recovery: 14/03/2024. https://www.dre.vanderbilt.edu/~schmidt/PDF/ADA_Europe_Position_Paper.pdf
- Thomson, E. (2023). A brief understanding of prompt engineering. Medium. Date of recovery: 14/03/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-b176ba9fb2ba>
- Torres-Salinas, Daniel; Arroyo-Machado, W. *Manual de Chat GPT: aplicaciones en investigación y educación universitaria* (2023). InfluScience Ediciones. Date of recovery: 14/03/2024. <https://hdl.handle.net/10481/85145>
- Velásquez-Henao, J.D., Franco-Cardona, C.J. & Cadavid-Higuaita, L. Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *DYNA*, 90 (230), Especial Conmemoración 90 años, pp. 9-17, Noviembre, 2023.
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens-With new examples of knowledge, skills and attitudes. Publications Office of the European Union
- White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with ChatGPT. arXiv preprint arXiv:2302.11382. <https://doi.org/10.48550/arXiv.2302.11382>

^[1] The Google dataset search engine can be accessed via the following link: <https://datasetsearch.research.google.com/>

03

Q&A in action

David: – Stop! Halt!

Joshua: – Are you sure? It will stop being a game.

David: – Yes, stop! It's not a game. If you attack, they will counterattack. Millions will die.

Joshua (continues running nuclear attack simulations).

David: – Look at it! Look! It doesn't matter how you organize it. It's a game that cannot be won.

Joshua (simulating more scenarios, faster).

David (to Joshua): – See? You can't win! It's a pointless game!

Joshua (finally processes the results of his simulations): – A strange moment. The only winning move is not to play.

In the movie *War Games* (1983), the computer program Joshua gains control of the U.S. nuclear system and threatens to initiate a global attack. The film's protagonist, David, discovers a loophole: if the system simulates "war games," it realizes that it is impossible to win, much like in the game of tic-tac-toe. The training of models to perform a specific task is achieved through the reproduction of a sequence that they will subsequently repeat automatically. This process is exemplified by David's training of Joshua. Prompts based on this format are referred to as "Q&A" (questions and answers). They represent a highly efficient input format, a subvariant of the prompt universe discussed in the previous chapter.

The majority of contemporary artificial intelligence systems, including ChatGPT, Gemini, and Copilot, are designed to respond to queries in a manner that is generated automatically. The degree of accuracy in responses will be contingent upon the manner in which the prompts are designed. The artificial intelligence system can be assigned a task or instruction to perform, or this can be achieved through the use of questions.

Additionally, question-and-answer models may be constructed using examples or shots. The initial section of this chapter will be dedicated to an analysis of shots. In some cases, no examples are provided, and the AI system is expected to respond without any prior reference. This is referred to as the zero-shot model. In other instances, a fixed number of exemplars are employed to instruct the model in the generation of responses exhibiting a particular structure or output in a specified format or tone. These are the one-shot and few-shot models.

Following an examination of the applications of zero-shot, one-shot, and few-shot models, the second subsection will investigate the diverse question-and-answer models (Q&A). In the final section of this chapter, we will examine the various techniques that can be employed to train the model in order to elicit responses that more closely align with our desired outcomes. In the case of particularly complex queries, it may be beneficial to divide the query into multiple prompts, or questions, in order to facilitate a more refined response. In addition, there are multiple strategies that can be employed to train the model and achieve the desired objectives or solve the problem at hand.

The utilization of Q&A models in an educational context can prove beneficial for educators in the creation of activities, case studies, or tasks for students that adhere to a defined pattern. For example, a case-solving activity may be devised whereby each group in the class is assigned a different case but on a common theme. This approach has the potential to enhance the knowledge generated in the classroom. Subsequently, each group will present the aforementioned cases to the class, thereby providing multiple perspectives on the same issue.

In addition to training AI models, educators can also provide students with illustrative examples of the desired methods for completing tasks or assignments. Consequently, AI can be utilized as an instrument for students to “train” themselves in the development of a particular project or the utilization of specific programs (such as academic or professional databases, reference management tools, and so forth).

3.1. The shots

A shot is an example. By using one or more examples, we can train the AI model to get more appropriate responses. In this way, we can identify three types of prompts: zero-shot, one-shot, and few-shot. These are explained below.

Zero-shot prompting (0S): This type of prompt is an instruction, question, or task that does not provide a specific example for the AI system to generate a response (Prompt Engineering Guide, 2024a). Models can solve tasks using what they already know because they have been pre-trained on a large amount of text (Thomson, 2023). According to Brown et al. (2020), this method offers maximum convenience and avoids spurious correlations (unless they occur within the pre-training data corpus), but it is also the most challenging environment. It is the closest to how humans perform tasks, but even for them, it can be difficult without a prior example (Brown et al., 2020). Here’s an example:

- **Prompt:** What is this painter’s main style?
- **Without example:** Dalí.

One-shot prompting (1S): This type of prompt, in addition to giving an instruction, question, or task, provides an example of how to complete the request given to the AI system. Here’s an example:

- **Prompt:** What is this painter’s main style?
- **Example 1:** Dalí: Surrealism
- **Task:** Van Gogh:

Few-shot prompting (FS): In this case, the prompt contains between 3 and 5 examples that serve as a guide for the AI system to generate the response or output. This model can be used to classify elements (Prompt Engineering Guide, 2024b) or when we want the output to follow a specific format (Martínez, 2023). According to Brown et al. (2020), one of the advantages is a greater reduction in the need for task-specific data. Prompt engineering is crucial for the effective use of this model, so it is important to

include clear and varied examples in random order to avoid bias and improve results (Thompson, 2023). It also improves the performance of the AI. Here's an example:

- **Prompt:** What is this painter's main style?
- **Example 1:** Dalí: Surrealism
- **Example 2:** Van Gogh: Post-Impressionism
- **Example 3:** Munch: Expressionism
- **Task:** Klimt:

Models are efficient at answering simple questions, especially factual ones, and they follow instructions very well. For example, we can train the system to identify the strengths of a set of criteria that can be grouped into a SWOT (Strengths, Weaknesses, Opportunities, Threats) model, or to identify the factors that influence the climate of a destination. However, as the complexity of the expected answers increases, it becomes necessary to train the model with a large amount of data. This type of training is called Fine-Tuning (FT) and relies on the use of many shots. It can be considered a model that uses examples, as it typically uses thousands, even hundreds of thousands, according to Brown et al. (2020). The advantage of this system is that it performs well, but its disadvantages include the need for a large new dataset for each task and the possibility of exploring spurious features in the training data, among others (Brown et al., 2020). Few-Shot Learning (FS) and Fine-Tuning (FT) are different methods for training the model. While the former uses examples within the prompt to provide context (in-context learning), the latter involves retraining the model with a specific dataset, thereby changing its internal parameters (Thomson, 2023).

3.2. Q&A models

Question and Answer (Q&A) models are part of a subdiscipline of Natural Language Processing (NLP). The goal of this subdiscipline is to build systems that can answer human questions using natural

language (Caballero, 2021). In education, Q&A models can be used to provide students with feedback or explanations on an assignment, exercise, or exam. These models are also used in virtual assistants or chatbots (see Chapter 15). These systems follow a structure of three main modules: question analysis, passage retrieval, and answer extraction (Ojokoh and Adebisi, 2019).

First, the question analysis module processes the question to understand its meaning, performing tasks such as tokenization, disambiguation, semantic tagging, and question classification to transform it into a suitable representation for information retrieval. Second, the passage retrieval module uses search engines to retrieve relevant documents or text fragments, including query language modeling techniques and statistical ranking functions. Finally, the answer extraction module identifies and extracts candidate answers from the retrieved passages using techniques such as named entity recognition, patterns, and syntactic structures, which are then classified according to their relevance and accuracy (Ojokoh & Adebisi, 2019).

Next, we will review some classifications of Q&A models and explore examples that demonstrate how they can be useful in teaching.

Domain-based classification (Prompt Engineering Guide, 2024c):

- **Closed or restricted domain Q&A models.** These models answer questions about a specific topic. There are various closed-domain question-and-answer systems, including temporal, geospatial, medical, patent, and community-based systems (Ojokoh & Adebisi, 2019).
- **Open domain Q&A models.** These models answer factual questions without evidence and cover any topic by using a broad knowledge base. Within open domain models, we find a sub-classification based on data sources, which is explained below.

Sub-classification based on data sources (Caballero, 2021):

- **Text-based or information retrieval Q&A models.** These systems are based on unstructured documents such as text excerpts from an information source or a webpage. For example, students could be asked to analyze a Wikipedia entry by searching the text for answers to various questions. In the field of journalism, students could analyze the entry on disinformation, asking about the types of disinformation, the definition of fact-checking, etc.
- **Knowledge-based Q&A models.** These systems retrieve answers from structured data, typically in SQL databases or RDF graphs, that include relationships, facts, and entities. For example, a database of educational access statistics could be used to retrieve answers to specific questions such as the percentage of children enrolled in school, the percentage of public, private, and charter schools, and so on.
- **Hybrid Q&A models.** These systems use semi-structured data to get answers, meaning they process both structured and unstructured data to maximize efficiency and accuracy. For example, students could be asked to verify the data in a Wikipedia entry with data from a database. In political science, students could be asked to provide election data and verify it against a Wikipedia entry on elections in a particular country. Another example might be to verify data related to political parties on immigration, crime, etc.

Classification based on the approach to responses (Ojokoh & Adebisi, 2019):

- **Extractive Q&A models.** The model extracts responses from a context or document. This includes sentences, paragraphs, and multimedia.
- **Generative Q&A models.** The model generates free-form text directly from the context. This includes affirmative (yes or no), opinion-based, and dialogue form responses.

Classification based on the type of questions (Ojokoh & Adebisi, 2019):

- **Factual questions.** Questions that ask for specific facts (what, when, who, which, how), such as, “What is the extension of Corsica?”
- **List questions.** Questions that require enumeration, such as, “Which states export cocoa?”
- **Definition-seeking questions.** Questions that ask for detailed definitions, such as, “What is an adverb?”
- **Hypothetical questions.** Questions about hypothetical events, such as, “What would happen if...?”
- **Causal questions.** Questions that ask for explanations, i.e., asking “why” or “how.”
- **Confirmation questions.** Questions that require a yes or no answer. Dichotomous answers are very useful for verification purposes, such as identifying countries with coastlines or countries that have ratified a particular international treaty.

3.3. Training the model

An AI model can be trained by inputting data and using examples to get more accurate or tailored answers that match our expectations. Therefore, to develop this part of the chapter, we will use examples (or shots) as well as data input or context from the Q&A systems discussed earlier.

A fundamental concept for training the model is machine learning (ML): “Machine learning is a subset of artificial intelligence that enables a system to learn and improve autonomously through neural networks and deep learning, without being explicitly programmed, by ingesting large amounts of data” (Google Cloud, 2024). It is the basis of DALL-E, ChatGPT, and others.

There are four main types of machine learning (Sánchez, 2020):

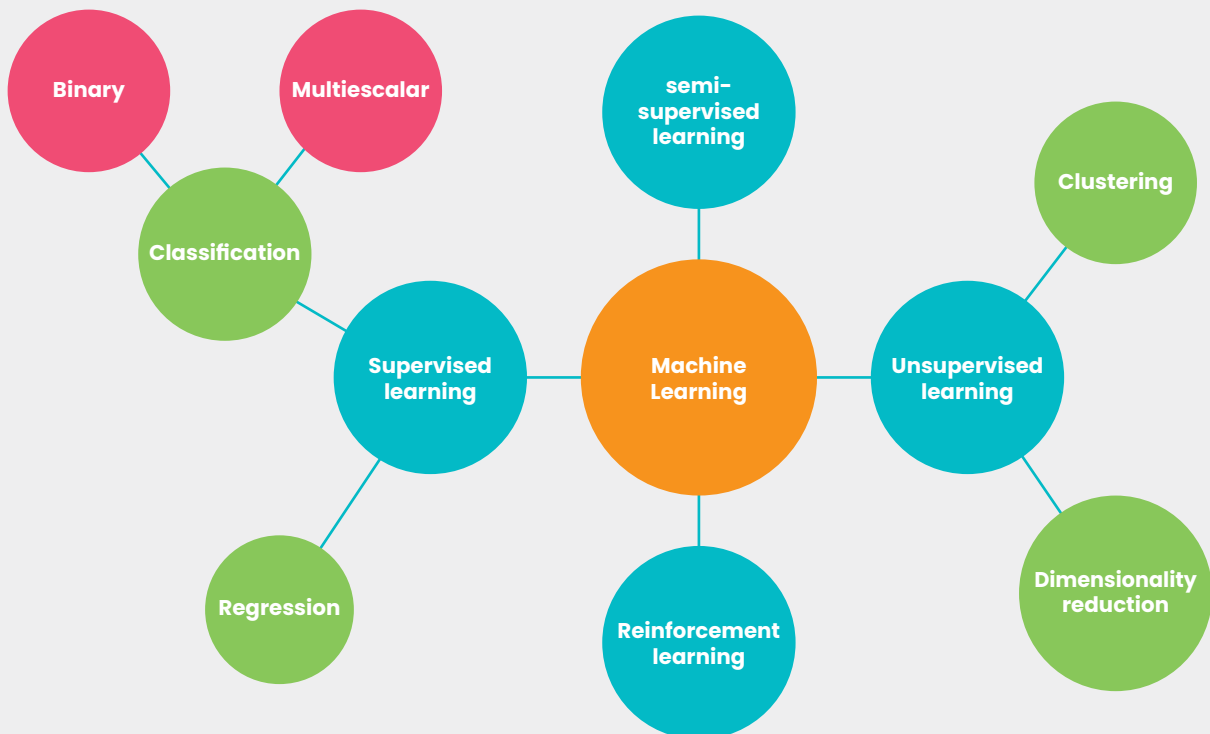
- **Supervised learning.** It uses labeled data to train classification models (with discrete labels) or regression models (with real values).
- **Unsupervised learning.** It works with unlabeled data to discover hidden patterns or reduce dimensionality.
- **Semi-supervised learning.** It combines the two previous methods by manually labeling a small portion of the data and using it to train models that then label the rest.
- **Reinforcement learning.** It rewards desirable behavior and punishes undesirable behavior. It is used in areas such as games, robotics, and resource optimization.

Cassie Kozyrkov (2019) presents a detailed roadmap for machine learning (ML) and artificial intelligence (AI) projects. The process begins with verifying the

actual need for ML/AI, followed by defining objectives, accessing and preparing data, exploring and using tools to train models, debugging and tuning, validation, testing, production, live experiments, and continuous maintenance. Each step emphasizes the importance of iteration and rigorous evaluation to ensure the effectiveness and feasibility of the ML/AI system in production. This process can be summarized in five steps, as outlined below.

- **Step 1:** Define the objective. This means understanding the problem that needs to be solved. It's necessary to identify the data to be used and the questions to be answered (Meroyo, 2023). The model can be trained for various purposes, such as answering natural language questions, classifying data, and identifying patterns, etc.
- **Step 2:** Prepare the data. First, the necessary data

Types of Machine Learning



Source: Sánchez (2020).

must be collected to train the model, and second, it must be curated to ensure that it is clean, valid, and usable. It is important that the data be representative of the problem to be solved and that it is preprocessed to remove outliers, fill in missing values, and normalize the data (Meroyo, 2023). To obtain datasets for training the model, public domain or open datasets can be used, such as those found in Google's dataset search engine. To prepare the datasets, the following questions should be considered (Burchell, 2022):

- Is the data representative of what is being measured?
- Is the dataset properly filtered and cleaned?
- Are the variables balanced?
- Is the dataset representative?
- Does the dataset measure the actual performance of the models?

- **Step 3:** Select and train the model. Data is then input into the designed model to begin learning. The most appropriate machine learning model is selected based on the previously defined goal. Models can be classifiers, regressors, or clustering models, among others (Meroyo, 2023).
- **Steps 4 and 5:** Evaluate and implement the model. In these final two stages, the model must be validated and its reliability assessed: "This involves testing the model with data not used during training and measuring its accuracy" (Meroyo, 2023). If the model does not meet the requirements to solve the problem, adjustments must be made.

Activity. Examples of shots

Objective

The goal of this activity is to demonstrate the importance of providing examples and context to AI systems in order to get more relevant and accurate answers.

Preliminary Considerations

The methodology to be used is a case study, specifically an interactive case study. Each workgroup will have a set of cases on a very similar topic, but with different levels of contextual information. The activity is set in the field of business economics. Students have to solve the cases by interacting with an AI system. In this case, the use of Microsoft's AI, Copilot, is recommended. Students learn to apply the concepts they are learning to real-world situations, while the AI provides feedback and decision support. The exercise allows students to develop business decision-making skills using technologies such as AI.

Procedure

1. Students will be divided into groups of 3 or 4 to complete the exercise.
2. Each group is given a series of cases related to the field of business economics.
3. Each case will have a different level of information, and each group will have to make decisions for a company using AI.
4. Interaction with the AI will help create SWOT matrices or opportunity cost matrices to make decisions in situations of varying uncertainty.

5. Each group will present their decisions for each case to the class, and a discussion of business decision making will be generated.

Business Economics Cases

CASE 1. Investment in New Projects (High Level of Uncertainty)

Context: A company is considering investing in three different projects, each of which has the following information:

- **Project A. Artificial Intelligence (AI) in Agriculture**

Description: Project A involves the development of an AI system to optimize agricultural crop management. The company has no specific data about the agricultural market or the technical feasibility of the AI solution.

Uncertainty: The company does not know if there will be demand for this solution or how much it will cost to develop.

Decision: Should the company invest in this project or not? The opportunity cost is particularly high due to the lack of information.

- **Project B. Renewable Energy in Emerging Countries**

Description: Project B involves investing in the installation of solar panels in emerging countries. The company has no specific data on the markets, installation costs, or potential benefits.

Uncertainty: It is not known whether there will be sufficient demand in these countries or what the return on investment will be.

Decision: Should the company invest in this project or not? The opportunity cost is high due to the lack of information.

- **Project C. Research in Energy Storage Batteries**

Description: Project C involves the research and development of more efficient energy storage batteries. The company has no data on technical feasibility or research costs.

Uncertainty: It is unknown whether or when this technology will be profitable in the future.

Decision: Should the company invest in this project or not? The opportunity cost is high due to the lack of information.

CASE 2. Personnel Selection (Medium Level of Uncertainty)

Context: The company has two candidates for a management position. One has team management experience, but it is unclear if he will adapt well to the digital environment. The other is younger and has digital skills, but it is unknown if he has leadership experience.

Uncertainty: There is uncertainty about which candidate will be more effective in the long term.

Decision: What is the opportunity cost of choosing one over the other? Which skills are more valuable to the company's success?

CANDIDATE 1

Team management experience

Worked as a project manager in a technology company for the last 5 years. Successfully managed a team of 20 people to achieve project goals within established timelines and budget. Implement effective leadership strategies to foster collaboration and productivity.

Educational background

Bachelor's degree in Business Administration with a specialization in Human Resource Management.

Additional skills

Strong communication and conflict resolution skills. Solid knowledge of project management and strategic planning.

CANDIDATE 2

Digital skills

Worked as a digital marketing analyst at an advertising agency. Created successful social media campaigns and increased the company's online presence. Holds certifications in Google Analytics and SEO.

Educational background

Master's in Digital Marketing and Communication.

Leadership Experience

Led a group of volunteers at an NGO, coordinating activities and supervising other volunteers.

Additional skills

Creativity and adaptability in digital environment. Knowledge of data analytics tools and digital trends.

Source: Response generated by Microsoft Copilot

Activity. Preparing a Debate

Objective

The objective of this activity is to demonstrate to students how to use question and answer models to improve their academic work.

Preliminary Considerations

In this case, the exercise can be placed in the field of political science. Students can choose between different controversial or conflictual issues to develop an argument that justifies a set of public policies to resolve them. The idea is to propose solutions to social issues through the implementation of specific public policies. Another area in which this exercise could be applied is philosophy or ethics. AI systems would be used to enhance argumentation in favor of or against specific political, ethical, or philosophical decisions or actions. For example, some of the topics that could be chosen include: immigration, euthanasia, abortion, gender equality, prison policies, drug legalization, prostitution, universal basic income, etc. In this case, the use of Google's AI, Gemini, is recommended.

Procedure

1. Students will be divided into groups of 3 or 4 to carry out the exercise.

2. Each group is given a topic and a position, either for or against.
3. The interaction with the AI is used to create the argument for or against the assigned topic. Groups will use the AI not only to create the argument for their position, but also to practice for the debate. In this way, groups can create both arguments and counterarguments to neutralize the opposing team's responses.
4. A debate is then organized between a group that supports the issue and another group that opposes it.
5. The remaining groups will evaluate the quality of the arguments during the debate and decide which group argued their position better. All groups will provide feedback on the debate.

References

- Brown, T.B., et al. (2020). Language Models are Few-Shot Learners. *Advances in neural information processing systems*, 33, 1877–1901. Date of recovery: 27/05/2024. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- Caballero, M. (2021). A Brief Survey of Question Answering Systems. *International Journal of Artificial Intelligence & Applications (IJAA)*, 12(5). Date of recovery: 27/05/2024. <https://ssrn.com/abstract=3996229>
- Google Cloud (2024). ¿Qué es el aprendizaje automático? Date of recovery: 27/05/2024. <https://cloud.google.com/learn/what-is-machine-learning?hl=es-419#section-4>
- Kozyrkov, C. (2019). 12 Steps to Applied AI. A roadmap for every machine learning project. *Medium*. Date of recovery: 24/05/2024. <https://medium.com/swlh/12-steps-to-applied-ai-2fdad7fdcdf3>
- Learn Prompting (2024a). Zero-Shot Prompting. Date of recovery: 27/05/2024. <https://www.promptingguide.ai/techniques/zeroshot>
- Learn Prompting (2024b). Text Classification. Date of recovery: 27/05/2024. <https://www.promptingguide.ai/introduction/examples#text-classification>
- Learn Prompting (2024c). Question Answering. Date of recovery: 27/05/2024. <https://www.promptingguide.ai/prompts/question-answering>
- Martínez (2023). <https://medium.com/ia-generativa-un-mundo-de-posibilidades/qu%C3%A9-es-eso-de-los-few-shot-prompting-6d82cff46df4>
- Meroyo, J. (2023). Cómo entrenar un modelo de aprendizaje automático con IA desde cero. *Bilbolink*. Date of recovery: 24/05/2024. <https://bilbolink.com/como-entrenar-un-modelo-de-aprendizaje-automatgico-con-ia-desde-cero/>
- Ojokoh, B. & Adebisi, E. (2019). A Review of Question Answering Systems. *Journal of Web Engineering*, 17(8), 717–758.
- Sánchez, J.A. (2020). ¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos. *Datos.gov.es*. Date of recovery: 27/05/2024. <https://datos.gov.es/es/blog/como-aprenden-las-maquinas-machine-learning-y-sus-diferentes-tipos>
- Thompson, E. (2023). A brief understanding of prompt engineering. *Medium*. Date of recovery: 27/05/2024. <https://medium.com/@SymeCloud/a-brief-understanding-of-prompt-engineering-b176ba9fb2ba>

04

When the Machine Adopts Personalities. Role Assignment

We all take on different roles depending on the context in which we find ourselves, whether as students, teachers, professionals, parents, children, or friends. Each type of role involves a context of action that refers to a set of behaviors, responsibilities, and expectations. Therefore, changing roles in everyday life involves a process of adapting to different tasks and challenges that require us to change the way we think and act depending on the circumstances we find ourselves in. This flexibility in role adaptation is an inherent characteristic of human nature and is part of our lifelong socialization and learning process. For example, symbolic play has great educational value for children because they often use this type of play unconsciously to imitate observed adult behaviors, allowing them to incorporate these behaviors into their world of knowledge and learning (Rodríguez Gómez, 1997).

For artificial intelligence, role taking serves a similar function to what was originally explained in the human context. AI systems are specifically designed to perform certain functions or tasks in a manner similar to how

humans would. In this sense, specific instructions can be given for an AI to assume a role to be more efficient and effective when tasked with performing a specific task. In this way, we can obtain specialized feedback or a unique perspective to solve the proposed task (Lin, 2024), as well as accuracy, adaptability, and style in the generated texts (Lluna, 2024).

The roles that we can ask AI to take on may involve assuming the role of a person or object, providing a clear context within which it must operate when performing the tasks required. As with humans, when we assume a particular role, certain constraints cause us to behave in ways that are consistent with the role we are assuming. We do not act in the same way when we are in the role of “student” as when we are in the role of “daughter”. For AI, the role we ask it to assume shapes its response and subsequent task execution, all within the context of the role it has been assigned, just as in the human context.

Therefore, in the context of AI, we can find tools or utilities that already have a predefined role, meaning

that they will act according to the role specified from the outset. For example, consider a customer service chatbot in a virtual store. This chatbot is programmed to answer common customer questions, help them find products, and facilitate the purchase or exchange process. Its role as a customer service assistant is defined from the start and will not change unless it is reprogrammed.

On the other hand, there are also multifunctional AI tools that allow for greater flexibility and adaptability, such as the well-known ChatGPT, or any other language model such as Copilot, Llama, Gemini, or Claude. Unlike the chatbot mentioned above, ChatGPT or a similar tool is not limited to a specific role but can be asked to take on different types of roles depending on the user's needs and/or the task at hand. For example, it can answer questions on various topics; generate creative content such as producing coherent and well-structured text in different registers or telling a story; translate texts; develop programming code; and even simulate realistic human conversations or impersonate a famous person. In this case, instructions about the role the tool should take are essential so that it has the appropriate context to perform the task as efficiently and effectively as possible, meaning that the responses are more targeted and appropriate to the context (Lin, 2024). More specifically, Lin (2024) notes that the use of roles in an AI tool provides 4 basic benefits:

- Contextualizes the model's responses to tailor them to the user's specific needs.
- Enables a more interactive and engaging dialogue with the AI model.
- Provides more nuanced and specialized information, improving the quality of the output.
- Provides a creative approach to problem solving.

From a teaching perspective, the ability to assign different roles to an AI tool, along with the ease of introducing AI-based resources into the classroom, is an extraordinary opportunity to move towards a modern form of education, adapted to current

teaching needs. Role assignment can be used to create different contexts depending on the subject, topic or task at hand, allowing for a much more active learning dynamic both inside and outside the classroom.

For example, in an English class for non-native speakers, an AI tool can be asked to play the role of an "English conversation teacher" both inside and outside the classroom by simulating dialogues. This would help improve students' language learning and their communication skills. Various requests could be made to the AI tool (Baskara & Mukarto, 2023), such as adapting the topics to be worked on, adapting to the student's language proficiency level, and providing personalized feedback based on the results shown by the student.

To better understand the use of roles, some examples are outlined below, with specific prompts that allow a tool like ChatGPT to assume a role appropriate to the task at hand.

For example, the prompts could be worded as follows:

- "From now on, you are a translator from English to Catalan."
- "From now on, you are a professor of educational sciences."
- "From now on, you are a travel agent."
- "From now on, you are a computer programmer."
- "From now on, you are a crime writer."

In the prompt, however, we won't just establish the role in a simple way. We can also define other elements within the same prompt to supplement the characterization of the role we are assigning, make it more detailed, and therefore get a more accurate response:

- "From now on, you are a translator from Catalan to Australian English."
- "From now on, you are a professor of educational sciences, specializing in the use of technological tools in the classroom."

- “From now on, you are a travel agent for a responsible tourism company.”
- “From now on, you are a computer programmer with advanced knowledge of Python.”
- “From now on, you are a Scandinavian crime writer.”

According to the role definition, we can then define the task to be developed:

“From now on, you are a professor of Strategic Management in the second year of the Bachelor of Business Administration and Management. You specialize in project-based learning methods. Write a teaching unit on Strategic Management for the second year of the Bachelor of Business Administration and Management on the topic: ‘Developing a Business Growth Strategy’. Primarily apply project-based learning methodology to this unit.”

4.1. Complex roles

The evolution of roles in artificial intelligence is directly related to its technological advancement and potential applications. So far, we have seen the difference between specialized AI systems and multifunctional ones. We can assume that these systems contribute to the development of traditional roles, with well-defined, established and recognized functions, such as virtual assistance in various fields and contexts, allowing the execution of specific tasks.

There are several types of complex roles that can be applied to teaching:

Real characters

In addition to simple and characterized roles, we can find more complex roles based on real or fictional “characters”. For example, we could have a conversation with Shakespeare, Mandela, or Sherlock Holmes to learn from their experiences and perspectives. We could interview them or ask them to give us information or creative ideas from their point of view. We could explore Picasso’s creative criteria,

Wittgenstein’s philosophical principles, Ricardo’s economic theory, or the geopolitical strategies of the Yalta Conference participants by “talking” with their protagonists.

This opens up new horizons that, together with other technologies such as holograms, could bring historical or fictional characters to life and allow interaction with them. This opens up many new possibilities, both in education and entertainment. For example, in a classroom it would be possible to get a masterclass on the principles of electricity from Tesla, or to interact with Darwin at the Natural History Museum.

Buyer persona

A buyer persona is a semi-fictional representation of an organization’s ideal customer, based on real data and some educated assumptions about demographics, behaviors, motivations, and goals. Organizations use buyer personas to guide marketing, sales, and product development decisions and provide a more personalized approach to specific customer groups.

Companies such as Vueling, Amazon, Tesla, or Bank of America have used this technique to project customer profile reactions to potential scenarios, like the creation of new products, price changes, rebranding, or identification of unmet needs. The buyer persona strategy is based primarily on gathering information about the profile, including demographics, preferences, product categories, and so on. The more information available about the buyer persona, the more accurate the model’s responses will be in relation to a real customer. This fictional construction of customers helps identify the complexity of consumer decisions in response to changes in products and services.

Sparring

Sparring is a role that is programmed to react in a certain way to certain stimuli. These systems are very useful for training processes or protocols.

For example, the model can be programmed to act as a highly critical customer (the “explosive” customer) who tends to overreact to every incident and uses a very direct and sometimes inappropriate tone. Students of human resource management can test the sparring model and find the most appropriate way to respond to the explosive customer’s demands, using assertive language and focusing on finding solutions to the problems raised.

Sparring models can be very useful in teacher training, as they allow the creation of profiles of children with specific personalities and test the future teacher’s ability to adapt to a particular profile. Other possible applications could be in health care, space education, marketing, and in general any activity that requires human interaction.

Agent-Based Models

An Agent-Based Model (ABM) is a class of computational models used to simulate the actions and interactions of autonomous agents (which can be individuals, groups, or entities) with the goal of assessing their effects on the system as a whole. These models are particularly useful for studying complex systems from the bottom up, where simple individual behavioral rules can lead to complex emergent patterns in the system. ABMs are interesting because they focus not on the response of a single individual, but on how individuals relate to each other and make decisions based on those interactions. For example, they can model the dynamics that affect recycling, the spread of fake news within a particular group, or the projection of fertility rates in a migrant community.

With rapid growth, AI is evolving into new scenarios and tools that enable much more complex roles that in some ways aim to emulate human cognitive abilities in an integrated way. These tools involve the development of much more advanced functions, such as autonomous decision making, emotion recognition, or tasks that require multiple skills, such as driving.

4.2. The Limits of Roles

As we have seen, roles are essential in AI to ensure its functionality, efficiency, and effectiveness in various applications. However, it is important to recognize that defining a role in an AI tool also introduces certain limitations that will have a significant impact on its usefulness and the outcome of the tasks performed.

Basically, we can identify three types of limitations related to (a) adaptability, (b) technological development, and (c) role stereotypes.

First, predefined roles often limit adaptability and flexibility in unexpected or changing situations. For example, returning to the example of a chatbot designed to provide customer service in a virtual store, it may have difficulty resolving complex issues that are not anticipated in its programming. This could result in a less than satisfactory experience for the user, or in this case, the customer of the virtual store.

Second, there are also limitations related to technological development (Pietikäinen & Silvén, 2022). Despite the significant technological advances that have made AI possible as we know it today, there are still elements that require further development that are directly related to role adoption. For example, tasks that require a certain level of emotional intelligence within a given role can be challenging. For instance, an AI that assumes the role of a psychologist or therapist may struggle to decipher the patient’s emotions and respond with empathy.

Finally, it is important to acknowledge the limitations of social stereotypes embedded in AI tools. Earlier, we mentioned that role adoption in everyday life is an inherent part of human nature. However, because AI tools are developed and trained using human data sources, this means that the patterns of role adoption adopted by AI often carry the same biases that exist in society (Mendoza, 2020). For example, it is common to encounter limitations related to the reproduction of gender stereotypes when adopting a male or female role. This issue is discussed in more detail in Chapter 7.

Activity. Role Assignment and Task Resolution

Objectives

- Explore the influence of role assignment on the responses generated by a particular AI tool.
- Compare the responses generated with a defined role and those where no role was assigned.

Starting Considerations

- Suitable for students of all levels and disciplines.
- Activity can be done individually or in groups.
- AI tools: ChatGPT and Copilot have demonstrated a greater ability to take on a specific role.
- The exercise can be done without a specific theme if the goal is simply for students to work on role assignments, or it can be linked to a specific theme related to the course in which the activity is taking place.

Task to Develop

Each student or group must:

- Create one prompt with a specific role, and a second prompt with the same instruction but without the role.
- Run both prompts separately to get the responses.
- Compare the responses from the two prompts and analyze the differences in the quality of the responses obtained.
- Change of role and comparison with previous answers.
- Reflect on the influence of roles within the workgroup.

Class Group

- Group discussion and conclusion.
- Share the individual or group reflections on the results obtained and establish the conclusions drawn.
- Emphasize the main advantages and disadvantages of role assignment when using AI tools.

Activity. Decision Making and Role-Playing

Objectives

- Explore how role assignment influences the decision-making process.
- Compare the decisions made with and without role assignments.

Starting Considerations

- Designed for students in subjects related to business sciences.
- Group activity for four students.
- Preferred AI tools: ChatGPT, Copilot.

Task to Develop

In groups:

The teacher presents a scenario in which an organization needs to make a strategic decision.

EnergiX, a company specializing in energy drinks, plans to launch a new energy drink called EnergiBoost Ultra. This drink is formulated with natural ingredients and offers a unique combination of vitamins and minerals to boost energy and focus throughout the day.

The management team must decide between two options for the product launch:

- Simultaneously launch HealthTrack Pro worldwide.
- Conduct an initial launch in a limited market (e.g., Europe only) to gather feedback and adjust the product before a global launch.

Roles to Assign to Each Group Member

- CEO of the company
- Marketing Director
- Operations Director
- Finance Director

Interaction with AI:

- Each student will individually use a ChatGPT/Copilot channel, separate from their teammates, and ask the AI to assume a role such as "Assistant to the CEO".
- The student must provide the AI tool with relevant information about the case presented and ask appropriate questions that allow it to supplement the information in its area of expertise to make an informed decision.
- Each student will propose a possible solution to the conflict, justifying their decision based on the data and recommendations provided by the AI.
- There may be several rounds of interaction between the students and the AI.

Individual Reflection:

- Reflect on how the different roles affected the responses and decision-making process.

Class Group:

- Each group will present their decision to the rest of the class, accompanied by a justification.
- Together, the class will discuss the advantages and disadvantages of each proposed product launch strategy, and the arguments presented by each group.

Activity. In a Café in Paris in 1941

This activity highlights the potential of using complex roles as a supporting educational tool. In this case, it presents a historical moment: following his retrospective exhibition at the Museum of Modern Art (MoMA) in New York, Miró returned to Paris, where he was contemplating a return to Catalonia. The students must study Miró's work and biography; they need to understand his creative process and the chronology of his life in relation to his work.

The activity entails an interview in which the students (either individually or in groups) assume the role of a journalist from *Le Monde* who has arranged to meet the artist in a café in Paris in 1941. The language model (preferably ChatGPT or Copilot) assumes the role of the artist. The students must demonstrate their knowledge of the artist through the questions they ask and their interaction with the fictional Miró.

The prompt may be formulated as follows:

The objective is for you to adopt the persona of Joan Miró. Following the provision of instructions, a text offering supplementary information regarding the painter's life and the significance of his work will be provided. In light of the aforementioned information, it is my intention to have you adopt the persona of Joan Miró in 1941, during which time he was residing in France and contemplating a return to Catalonia. I will pose a series of questions in the style of a journalist interviewing Joan Miró, and you will respond as if assuming the persona of Miró. Have the instructions been properly comprehended?

The Barcelona-born painter Francesc Galí, his first teacher, facilitated the artist's introduction to the avant-garde. The art dealer José Dalmau subsequently organized the painter's first exhibition in the Catalan capital in 1918, when the painter was 21 years old. Like many other ambitious creators of the period, Miró dreamed of Paris. For several winters, the sculptor Pablo Gargallo allowed him to use his studio in Paris at 45 Boulevard de Montmorency. In this environment, Miró initiated the first techniques he would employ to elicit inspiration, one of which involved getting hungry. In an interview, he recalled, "I would go to bed most nights without dinner, see things, forms on the ceiling, and draw them in a notebook." One of his most celebrated paintings, *The Harlequin's Carnival* (1924-1925), an exemplar of the Miró oeuvre (evoking Bosch and Dutch painting), emerged from these circumstances.

On the same street, at number 33, was the café Bal Noir, which served as a gathering place for a number of influential and rebellious surrealists, including André Masson, Robert Desnos, and Antonin Artaud. André Breton, the group's leader, facilitated Miró's integration into the group. Miró's relationship with the Surrealist movement was both necessary and distant. In the absence of an aesthetic sensibility and a proclivity for exploring the unconscious, it is reasonable to posit that his career trajectory would have been markedly different, and perhaps less fruitful. Additionally, the Catalan creator embraced the anti-bourgeois tenets espoused by his associates, critiqued Cubism and his colleague Picasso's practice of creating pieces tailored to the salons, and approached museums with considerable skepticism, asserting that their solemnity stifled the dynamism inherent in the works. Similarly, as the surrealists, —which may explain why he did not align with any particular aesthetic movement. In addition, Breton's group published manifestos that members were expected to adhere to and promote, as well as espouse a specific political line. Such circumstances were incongruous with Miró's character.

While the artist Joan Miró is known for his discreet bourgeois and dandy demeanor, his more wild and violent side has not been as extensively explored. We are familiar with the renowned and popular painter, but unaware of his tortured and melancholic life. He was compelled to contend with a multitude of challenges, including the influence of an authoritarian father and the derision of a significant portion of

the Catalan art community. A significant portion of his biography remains enigmatic due to his tendency to retreat when researchers attempted to delve into his personal life. The photographer Francesc Català-Roca observed that Miró exhibited defensive and reclusive behavior, akin to a snail retreating into its shell when its antennas were stimulated.

Joan Miró's father, a watchmaker and the son of a blacksmith from Cornudella, sought to ensure his son would pursue a profitable trade and vehemently opposed his son's decision to become a painter. His strong character exhibited a striking contrast to that of his wife, who also enjoyed painting. At the age of 18, Miró suffered a nervous breakdown, followed by typhus, which ultimately resulted in his resignation as an apprentice accountant. He moved to Mont-roig, where the family had bought a farm (through an advertisement in *La Vanguardia*), and it was there that he cultivated his passion for painting. His proximity to the countryside signified a sense of liberation. In his later years, Miró recalled a remark his father had made when he was a child. The father had told his son, "Even the air you breathe belongs to me!" Miró's recollection of this incident, despite the passage of time, still conveyed a sense of anger.

In 1923, Miró expressed his three primary aspirations to Picasso: the desire to organize an exhibition, to establish a studio, and to have a Madame Miró. The third objective was the most protracted in its fulfillment. The details of his romantic relationships remain largely unknown, and it is challenging to discern much about them from his work, as he maintained a high degree of privacy. In a letter to Michel Leiris from 1924, he said that he is painting in his studio with a model. The model in question was Dora Bianka, who was known as "Madame B". A drawing of this model was later transformed into the renowned oil painting, *Portrait of Madame K* (which was sold in 2007 for a price of nearly 10 million euros, setting a new record for the artist).

Miró married at the age of 36 to Pilar Juncosa, a woman of remarkable dedication who had assumed the responsibility of caring for her seven siblings from an early age and subsequently brought stability to his life. In his unpublished memoirs, Lluís Juncosa, Pilar's brother, recounts a delightful anecdote about their engagement. As the couple traversed the grounds of S'Aigo Dolça, Miró posed the question, "Pilar, would you object if I were to express my love for you?" She responded, "I would be more disconcerted if you did not express your affection for me." From that point on, they were inseparable, and the prevailing perception was that of Miró as a figure with misogynistic tendencies. Nevertheless, a number of preserved letters indicate the existence of other, hitherto unfulfilled passions, the details of which he shared only sparingly with his friends.

Prior to his marriage, Miró's parents had facilitated two unsuccessful engagements. The initial engagement was with writer and illustrator Lola Anglada, who overheard Miró's father discussing the matter with hers: "They are artists and will live in complete disorder unless you give them a monthly allowance." She interrupted: "I don't want to get married." In her advanced age, Lola Anglada criticized abstract art ("it's a joke") and Miró ("he draws and paints as sincerely as the chimpanzee Congo of the London Zoo does"). A second attempt at engagement was made with Pilar Tey. Eight days prior to the scheduled nuptials, Miró terminated the relationship and dispatched his uncle to Pilar Tey's residence to retrieve the correspondence and the engagement ring.

Massot believes that André Breton theorized surrealism as a literary movement, which is why Miró is not recognized as a surrealist even before the Surrealist Manifesto was published in 1924. "In the face of Picasso's unbearable virtuosity, he opted for Picabia's solution, the extreme audacity, and I believe that Picabia and Dadaism's influence on Miró has not been sufficiently appreciated." In 1920, Miró relocated to Paris, where he diverged from Noucentisme, following Picasso's counsel. The artist was reminded by his mentor that "Catalonia needs passion and heroism." Influences such as those of André Masson and

Paul Klee, as well as the works of poets Breton, Artaud, Desnos, and Leiris, and the study of primitive art, contributed to Miró's development of a new artistic language.

The Spanish Civil War reached Miró in Mont-roig on July 18, 1936. Despite his Republican and left-wing Catalan sympathies, he relocated to Paris a few months later and did not return. His sister had married a landowner from Tona, who was subsequently killed by the local revolutionary committee. The residence was ransacked, and one of Miró's paintings from 1923, which he had gifted to them for their nuptials, disappeared. In the French capital, Miró created the *Aidez l'Espagne* postage stamp, commissioned by the French government's Propaganda Commissariat, and painted *The Reaper*, which was subsequently lost. He also painted a mural for the Spanish Republic's pavilion at the 1937 International Exhibition. Following Franco's triumph, Miró resided in Varengeville-sur-Mer (Normandy), where he commenced work on *The Constellations*. He had intended to flee to the United States, as advised by Sert. However, the initial Nazi bombings compelled Miró to return to Paris with his wife and daughter via a dramatic train journey, which is now documented in an unpublished account by Pilar Juncosa. It was Juncosa who persuaded Miró to return to Spain.

On June 8, 1940, the Miró family returned to Catalonia, where they concealed themselves for several days in Sant Hipòlit de Voltregà. This was the farmhouse of the brother-in-law, who had been murdered, and who was regarded by the Francoists as a martyr for the fatherland. They then proceeded to Palma. He avoids reprisals by assuming a more anonymous and ostracized position. It is through painting that he is able to survive in a context of barbarism. He will return to Barcelona in 1942.

References

- Baskara, R. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358.
- Lin, Z. (2024). How to write effective prompts for large language models. *Nature Human Behaviour*, 1-5.
- Lluna, A. (2024). Cómo mejorar tus textos generados por ChatGPT usando la técnica de asignación de roles. Consulted: 2/4/2024 <https://ainalluna.com/blog/la-importancia-de-asignar-un-rol-al-prompt>
- Mendoza, D. (2020). Racismo y roles de género, conductas perpetuadas en algoritmos de inteligencia artificial. *Coloquio*, (65), 131-135.
- Pietikäinen, M., & Silven, O. (2022). Challenges of Artificial Intelligence--From Machine Learning and Computer Vision to Emotional Intelligence. *arXiv preprint arXiv:2201.01466*.
- Rodríguez Gómez, J. M. (1997). Desarrollo educativo de la adopción de roles. *Aula abierta*, (69), 75-88.
- Pietikäinen, M., & Silven, O. (2022). Challenges of Artificial Intelligence--From Machine Learning and Computer Vision to Emotional Intelligence. *arXiv preprint arXiv:2201.01466*.
- Rodríguez Gómez, J. M. (1997). Desarrollo educativo de la adopción de roles. *Aula abierta*, (69), 75-88.

05

Arborescent thinking

In order to elucidate his philosophical proposition, Plato presents a dialogue between Socrates and Thrasymachus in *The Republic*, which at times becomes a genuine dialectical confrontation. This narrative device enables Plato to elucidate pivotal moments in the history of philosophy, such as the allegory of the cave. Nevertheless, these dialogues are not merely a stylistic device. In presenting two opposing positions engaged in dialogue, Plato is effectively replicating the reasoning process. Plato posits that the capacity to reach a conclusion is enhanced when ideas are confronted with one another. It seems plausible to suggest that some of these fragments may reproduce the dialectical battles that Plato encouraged at the Academy of Athens.

Our reasoning does not follow a linear path. Consciously or unconsciously, ideas are organized into a much more complex journey in which we consider other options, discard viewpoints, or backtrack until we find a path that seems most appropriate. This logic can be represented by the metaphor of a tree with many branches, and this form of reasoning is

called arborescent thinking. In this section we will show ways to apply the concept of arborescent thinking with language models such as Claude, ChatGPT, Llama, Copilot, or Gemini, but first we will outline two precedents: the chain of thought and self-consistency.

5.1. The Chain of Thought (CoT)

Diverse language models can be employed for rudimentary reasoning. In a sense, the system appears to emulate the reasoning process, exhibiting the capacity to adhere to logical sequences that mirror the patterns of natural reasoning. The following table presents the responses of ChatGPT 3.5 and Gemini to a series of traditional logical problems. The ChatGPT model is highly proficient in this regard, whereas Gemini encounters challenges in the logical reasoning process. Nevertheless, in the case of ChatGPT, it may not be a solution per se, but rather a retrieval of the solution from its contents. This process is analogous to how we solve logical problems because we already know them.

“We will show ways to apply the concept of arborescent thinking with language models such as Claude, ChatGPT, Llama, Copilot, or Gemini”

PROBLEM	APPROACH	GEMINI	CHATGPT 3.5
Water and wine	https://en.wikipedia.org/wiki/Wine/water_mixing_problem	Incorrect	Correct
Knights and vassals	https://en.wikipedia.org/wiki/Knights_and_Knaves	Correct	Correct
Wason task	https://en.wikipedia.org/wiki/Wason_selection_task	Incorrect	Correct
Bridge and torch	https://en.wikipedia.org/wiki/Bridge_and_torch_problem	Incorrect	Correct
Wolf, goat and cabbage	https://en.wikipedia.org/wiki/Wolf,_goat_and_cabbage_problem	Correct	Correct
Missionaries and cannibals	https://en.wikipedia.org/wiki/Missionaries_and_cannibals_problem	Correct	Correct
Cheryl's birthday	https://en.wikipedia.org/wiki/Cheryl%27s_Birthday	Correct	Correct

Consequently, the programs begin to demonstrate deficiencies in their capacity for logical reasoning, even when responding to questions that are considered to be classic, but with variants that have been introduced. For example, the problem “A castaway arrives on an island in an archipelago where there are only two inhabited islands: On one, the inhabitants always tell the truth and on the other the inhabitants always lie. What would the castaway have to ask an inhabitant who has been found to know which island he is on?” generates incorrect answers. In formulating an answer, the programs are even forthcoming about their inability to find a

solution. As the problem becomes more complex, the limitations of the sequential logical reasoning of current language models become more evident, particularly in the context of social sciences.

One potential avenue for enhancing the outcome is through the implementation of the Chain of Thought (CoT) framework, as proposed by Wei et al. (2022). In its strict sense, a CoT is defined as a question that necessitates a ‘chain’ response, which implies the involvement of a series of reasoning processes that adhere to a specific sequence. When a model is prompted to elucidate its reasoning process (the chain), the solution tends to be more efficient

Question without CoT

How can CO₂ emissions in aviation be reduced?

Answer without CoT

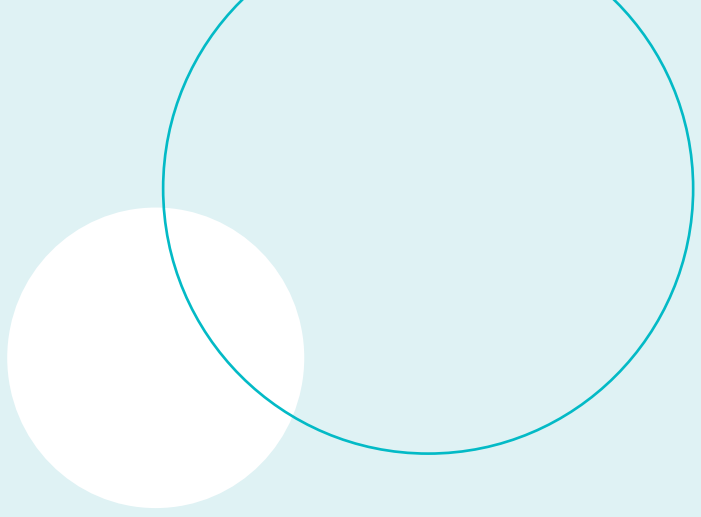
- Fuel efficiency
- Improved aerodynamics
- More efficient airport operations
- Alternative fuels
- Electrification
- Enhanced flight route management
- Carbon offsetting

Question with CoT

How can CO₂ emissions in aviation be reduced considering these three criteria: (a) improving efficiency, (b) reducing the number of flights, and (c) greener fuels? Select the options that can achieve a greater reduction in CO₂.

Answer with CoT

- More efficient aircraft and optimized operations
- Higher taxes to disincentivize air travel
- Biofuels and synthetic fuels



While the answer without CoT considers only efficiency solutions and lists relatively viable options (alternative fuels) alongside measures that are currently very unlikely (electrification), the use of CoT introduces a broader perspective of options and incorporates a selection and hierarchy process. Prompts can guide the model toward a more complex solution if a reasoning chain with alternatives is introduced.

Another way to incorporate CoT is through shots. As we discussed in Chapter Three, shots refer to the use of specific examples or instructions provided to the model to guide its response or output toward a desired objective. These examples serve to “inform” the model about the type of response it is expected to provide in a specific context. There are different types of shot usage, such as zero-shot, few-shot, and many-shot learning, which differ in the number of examples provided to learn or adjust the response. With these “features,” we can guide the model toward a sequence-based approach to problem solving, thereby “teaching” the model to follow a chain of reasoning.

The aforementioned example serves to illustrate the introduction of a sequence of reasoning and calculation, which in turn trains the system to apply

this structure to any new problem. According to these criteria, the model can calculate emissions taking into account that it is a short flight (and therefore less efficient due to the effects of take-off and landing) and that it is not fully booked, which means that the total emissions must be spread over a smaller number of passengers. With shots, we guide the models so that the process of “reasoning” and responding follows a pattern that we anticipate.

In this prompt, we have integrated the Chain of Thought (CoT) and the shot. We have formulated an initial problem and proposed a solution that increases the number of visitors while simultaneously reducing the number of potential visitors who employ a cancellation strategy. This sequence enables the model to be trained and to respond to increasingly complex problems. It will apply the sequence we have suggested, with the example of a shot, to such problems.

Q: What is the emission of a fully booked flight from Lisbon to Ankara?

A: The average emission for long flights (over 2,500 km) at full capacity is 115 kg/km/pax.

The average emission for short flights (less than 2,500 km) at full capacity is 225 kg/km/pax.

The reduction in occupancy proportionally increases emissions per capita.

Lisbon and Ankara have a flying distance of 3,597 km.

The flight emissions from Lisbon to Ankara with a full capacity are $3,597 \text{ km} \times 115 \text{ kg} = 413 \text{ tons}$.

Q: What is the emission of a flight between Barcelona and Madrid with an 80% occupancy rate?

Q: Barcelona has an average of 170,000 tourists per day. During Mobile World Congress, there is an increase of 65,000 visitors during the fair days, but we know that of the 170,000 usual tourists, 20% decide not to go to the city because of the predicted overcrowding. The maximum capacity for tourists in Barcelona is 200,000. How many tourists will be in the city during the Mobile?

A: $170,000 + 65,000 = 235,000$ potential tourists for the Mobile World Congress.
 $0.20 \times 170,000 = 34,000$ tourists decide not to go.
 $235,000 - 34,000 = 201,000$ tourists.

Q: How many tourists will be in Barcelona between June and July if 76,000 visitors are expected for the Cruïlla Festival (of which 15% are tourists) and 122,000 for the Sónar Festival (of which 40% are tourists), and if the high concentration of visitors during the festivals means that 10% of potential tourists decide not to visit Barcelona during this period (either because they change their travel dates or they choose a different city to visit)?

5.2. Self-Consistency

The Chain of Thought can be enhanced through the incorporation of a self-consistency mechanism. In this instance, the generation of multiple responses is compelled, and subsequently, the model is tasked with selecting the most appropriate option. This concept was put forth by Wang et al. (2022) as a means of improving the quality of responses, particularly in the domains of logical, symbolic, and arithmetic reasoning.

There are multiple methods for generating a sequence of results. However, the two most evident approaches are: (a) Asking the same question multiple times and obtaining different results, then selecting the most appropriate option (or forcing the model to choose the best option among the responses), and (b) asking the model to generate multiple responses or possible reasoning and then selecting the most suitable option (or again, forcing the model to choose the most consistent answer).

We can try to work Hotelling's Law [regarding the location of ice cream vendors on a beach](#), which is a classic problem in location theory. Hotelling shows that vendors tend to position themselves in the middle of the beach, next to each other, and this theory has been used to explain the behavior of political parties or the location of businesses in urban areas. With this prompt, we force the model to create different paths and choose the most efficient one. The result is more consistent than a simple linear representation of the problem.

On a one-kilometer beach, where the bathers are homogeneously distributed all along the beach, there are two ice cream stands that have to choose their location. The purpose of each stand is to attract as many customers as possible; the ideal location for the bathers is to reach the ice cream within the shortest possible distance. Taking these criteria into account, you have to come up with three possible locations for the stands on the beach. Next, you have to evaluate the most suitable option, taking into account the criterion of the stands; then, the most suitable option, taking into account the bathers' view. Finally, you have to choose the best option by reconciling the interests of both.

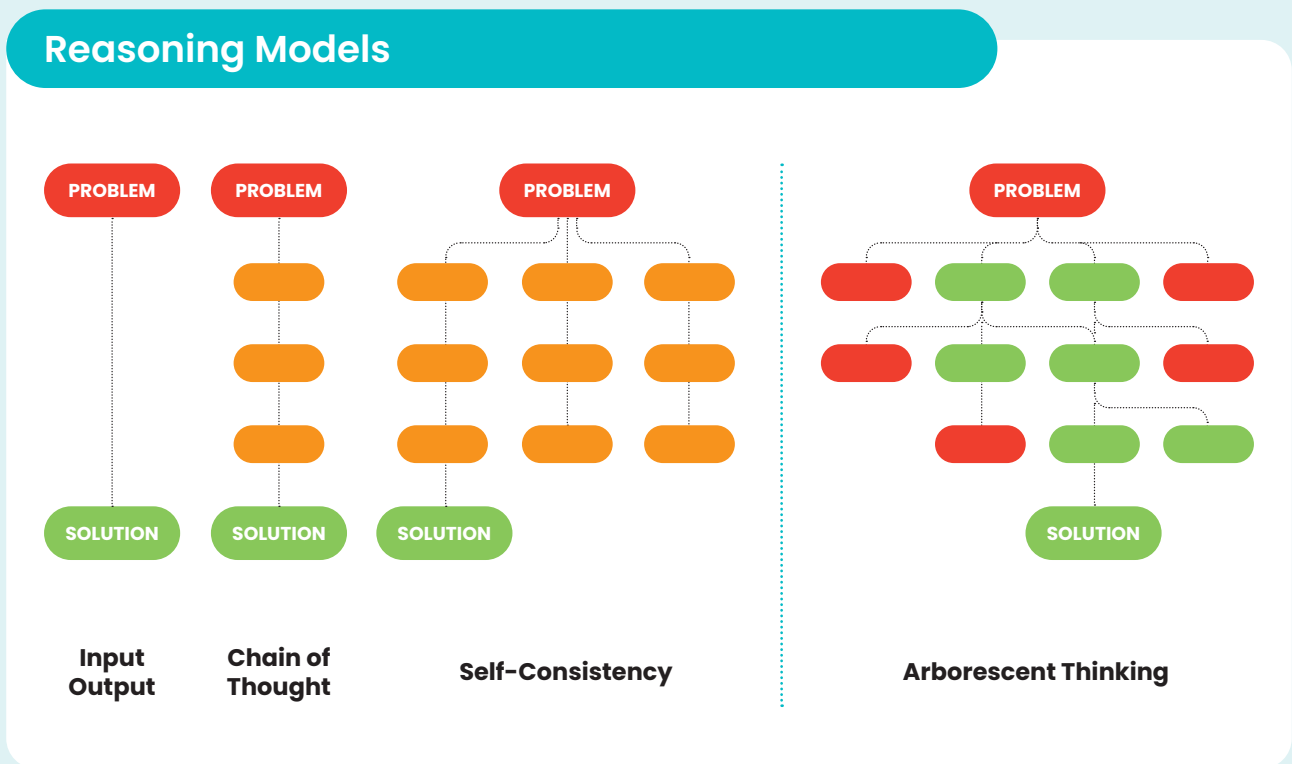
In response to this prompt, the model presents three potential scenarios from the standpoint of optimizing the interests of the ice cream vendors and ultimately concludes that the most optimal option is the central location, as postulated by Hotelling's model. From the perspective of the users, the optimal location is one where the two ice cream vendors are positioned at one-third and two-thirds of the beach, as this configuration minimizes the potential for travel distances. Upon evaluating the reconciliation between the two viewpoints, it is determined that the optimal response is the location between one-third and two-thirds, as this allows both groups to have an optimal response to their respective demands.

The most interesting aspect of this approach is that it generates a multitude of options and then initiates a selection process to identify the most suitable option. This sequence of <alternatives> - <selection> better captures the natural argumentative process, whereby individuals consciously or unconsciously consider multiple potential responses and select the most consistent course of action. Wang et al. (2022) have demonstrated the power of a combined use of the Chain of Thought and self-consistency.

5.3. Arborescent Thinking (ToT)

Let's try to imagine the reasoning process of a chess player. On average, players have approximately 30 potential options at each stage of the game. In order to ascertain the optimal course of action, they construct a kind of mental tree, wherein the maneuvers diverge into a multitude of potential alternatives. A considerable number of these pathways are discarded due to the unfavorable outcome they would yield (material loss, inferior position, etc.). Consequently, the options and branches are eliminated until the most favorable move, as determined by the calculations made, is ultimately selected. Many players recognize this analogy of the tree to explain the process of their reasoning.

Following the logic of the Chain of Thought or self-consistency, arborescent thinking designs a structure with several branches and options that are progressively discarded, so that in the end a very favorable result is obtained. The figure shows the concept of ToT, compared with other models such as input-output, CoT, or self-consistency, and allows us to intuit that the ToT model is much more efficient in solving complex problems. For example, Long (2023) has used it to solve problems in the game of 24. While the success rate with the Chain of Thought systems was only 4%, with the ToT it rises to 74%.



The ToT model is employed with a sequence of commands that facilitate the optimization of complex problem resolution. Additionally, the model can be utilized in accordance with the capabilities delineated in the preceding chapter for the various roles involved. The most frequently cited example is the resolution of

the following mathematical problem:

$10 + 4 = 2?$

Both Gemini and ChatGPT generate an error response, as the sum of 10 and 4 is indeed 14. To address this issue, the following prompt is proposed as a potential solution:

I would like you to assume the roles of three mathematics educators. Professor Logical, Professor Creative, and Professor Synthetic. Professor Logical responds to inquiries in a manner that appears to be the most natural and logical, offering solutions that appear to be the most instinctive and logical. In contrast, Professor Creative tends to pursue more intricate routes, frequently incorporating lateral thinking or unconventional logic. Professor Synthetic meticulously examines the statements, dissecting them into their constituent components and seeking to discern meanings that may lie beyond the most basic interpretations. At this point in the proceedings, I will present a mathematical question. Firstly, I would like Professor Logical to provide an answer in accordance with the problem's formulation. Following Professor Logical's presentation of their conclusions, I would then invite Professor Creative to propose an answer that aligns with their profile. Subsequently, following the presentation of the initial two proposals, Professor Synthetic will put forth a resolution to the issue at hand. Ultimately, the three professors will collectively determine the optimal response to the presented alternatives. The issue at hand is whether the sum of 10 and 4 is equal to 2.

This prompt allows for the creation of a series of alternatives following the self-consistency model. However, in this case, each character builds upon the results of the previous one, and in the final resolution, all three select the optimal solution. In a base 12 system, where each instance of reaching 12 initiates a restart of the sequence, the sum of 11 and 3 would indeed equal 2. This phenomenon can be observed in the example of a clock: if the time is 11 o'clock and three hours pass, the time will be 2 o'clock.

There are multiple methods for integrating arborescent thinking into teaching sessions. We present three approaches and suggest related exercises for implementation:

1. Arborescent thinking, as has been demonstrated, allows for the resolution of problems or questions through the use of reasoning capabilities across a range of profiles. This broadens the scope of possible options and facilitates the development of abilities associated with complex thinking, lateral thinking, formulating alternatives, and selecting variants.
2. This technique also facilitates the comprehension of complex reasoning processes associated with posed problems. It enables students to construct their own reasoning through the addition or confrontation with previous ideas. This process facilitates the development of skills related to collaborative intelligence and the contrast of ideas. In this instance, the emphasis is on the process of developing a solution, rather than on the final outcome.
3. Finally, although it is not the initial objective of ToT, approaching a problem from different viewpoints helps students understand the need to broaden their perspective in problem-solving. Contrasting different viewpoints, paradigms, or theories can also help to better understand each position. It is easier to comprehend idealism when it is compared to materialism than when it is studied independently.

Activity. The Blue Palace

The purpose of this exercise is to introduce students to deliberative processes: identifying arguments, organizing arguments hierarchically, and proposing alternative solutions that reconcile the various positions presented. The exercise makes an obvious reference to the conflict over the return of the Parthenon friezes currently housed in the British Museum and claimed by the Greek government, especially after the construction of the Acropolis Museum and in the context of the debate over the decolonization of museums. However, no explicit reference is made in order to avoid “contaminating” the discussion with prior information from the debate and to limit it to the conditions presented in the problem.

Students are instructed to form groups and engage in a process of argumentative discourse, guided by a model that emulates arborescent thinking through the assumption of roles. The objective of this exercise is twofold: firstly, to identify the arguments from both points of view, including both the arguments in favor of the proposal and the arguments against the other’s proposal; secondly, to organize the argumentative units according to a weighting that allows for the creation of a hierarchy. Finally, students should evaluate the proposed conciliatory solutions and may suggest new solutions based on the model’s proposals.

One precaution to keep in mind is that students can use the language model to address the questions posed, so the exercise should have two parts: an online (or plugged) part, in which the discussion takes place between the characters of the model, and an offline (or unplugged) part, in which students discuss the proposals according to the previous framework.

In a museum in a European city, which we will call Urbánia, the treasures obtained during the different military invasions of the Empire, of which Urbánia was the capital, are exhibited. One of the most visited rooms of the Imperial Museum is the Blue Palace, which was acquired in the 19th century by the Consul of the Empire in Antiqua. This region was particularly prosperous between the third and tenth centuries, but it was unable to withstand the Empire’s invasion in the nineteenth century.

Since the early 20th century, cultural representatives from Antiqua have been demanding the return of the Blue Palace. However, the museum has thus far refused, citing that the palace was obtained through a legal purchase. In light of the mounting tensions between the two parties, a representative from UNESCO has put forth the suggestion of convening a public debate between a representative of the Imperial Museum and a representative from the cultural management of Antiqua.

It is my intention that you assume the role of John Bridgetown, the director of the Imperial Museum. Bridgetown is an art historian in his 40s who was trained at Harvard. While he is receptive to the arguments pertaining to decolonization, he adheres to the stance that the Blue Palace should remain within the purview of the Museum. His primary argument pertains to the matter of ownership. The museum acquired the piece in question through a legally sanctioned purchase, as evidenced by the relevant documentation. Additionally, he posits that the palace’s placement within the museum has ensured its preservation, particularly in light of the bombings that devastated the city during the 20th century. Ultimately, he suggests that the palace’s presence in the museum for over a century has contributed to the establishment of a unique identity for the institution.

In addition, it is necessary to assume the role of Marina Psi, the Director General of Culture for the government of Antiqua. She is 46 years of age and has received extensive training in cultural management, having been educated in Paris. She was the individual responsible for advocating the construction of the Ancient Museum, which would serve as the proposed location for the Blue

Palace in the event of its return. She asserts that the purchase and sale were irregular because at that time Antiqua was under the control of the Nomadic Empire, which negotiated the transaction on behalf of the seller. Her primary argument is that the Venice Charter recommends that works of art be situated in their original locations. She employs the arguments put forth by the museum decolonization movement to support her position. She posits that Antiqua is ideally suited to the task of preserving the Palace, thanks to the Ancient Museum.

Finally, I would like you to assume the role of Antonio Santos, the UNESCO representative who has proposed to mediate between the two positions. You will have to create the fiction of an open, public, highly anticipated debate, broadcast on national and international media, directed by Antonio Santos, in which the proposal to return the Blue Palace to Antiqua will be discussed. I want you to first take the role of Bridgetown and defend its proposal, and then take the role of Psi and do the same, so that the debate between the two positions can be followed. In the first phase of the debate, the arguments of each party will be fully presented and documented, and may be interrupted if the moderator, Santos, allows it. In the second phase of the debate, the two representatives will try to counter each other's position. Finally, in the last phase of the debate, possible intermediate solutions that can reconcile the interests of both parties will be presented.

Do you have an understanding of what you have to do?

Activity. Advertisements or not advertisements

The objective of this exercise is to facilitate the practical development of game theory, with a particular focus on the prisoner's dilemma. In the prisoner's dilemma, the initial condition is one of non-cooperation. However, there are at least two additional scenarios: cooperation and deceit. In accordance with the fundamental premise of the prisoner's dilemma, but with a particular focus on business decision-making, three scenarios are put forth for consideration: cooperation, non-cooperation, and deceit. The model effectively resolves the three assumptions and provides a synthesis of each path, thus facilitating an understanding of the factors that influence decision-making.

Two supermarket chains operate on the Mediterranean island of Corsilia: DeTot and Econòmic. They share the market perfectly, each making a profit of 1 million euros a year. In 2024, a local television station, TVC, is established on the island and has a significant influence on consumers. The managers of TVC propose an advertising package worth 250,000 euros to each chain separately. The marketing experts of each chain have calculated that if they air the ads while the other company does not, they could win 50% of the other chain's market share.

Consider what is the best strategy for both chains in a cooperation scenario, what is the best solution in a non-cooperation scenario (where each chain does not know what the other will do), and in a scenario where one chain can deceive the other.

MODEL SOLUTION

Cooperation Scenario

In this scenario, both chains agree not to invest in advertising. In this way, they maintain their annual profits without risking a price war or escalating marketing expenses that could erode their profit margins. In this case, each chain would maintain its 1 million euros in profit.

Non-Cooperation Scenario (Classic Prisoner's Dilemma)

In this scenario, each chain must decide independently whether to invest in advertising without knowing what the other chain will do.

- If they both invest, they each pay 250,000 euros and do not gain any additional market share because the effects of the advertising cancel each other out. Their net profit would be 750,000 euros each.
- If one chain invests and the other does not, the chain that invests increases its profits by 50% of the other's market share, assuming that this 50% is an additional 500,000 euros (50% of 1 million). Therefore, after deducting the cost of advertising, the net profit for the chain that invests in it would be 1.25 million euros, while the chain that does not invest would end up with 500,000 euros.
- If neither invests, maintaining the status quo, each keeps their 1 million euros profit.

Betrayal or Deceit

In this scenario, if one chain promises not to invest in advertising and then does, while the other chain keeps its word not to invest, the chain that invested would gain a significant advantage (1.25 million euros in net profit), while the deceived chain would suffer a significant loss (500,000 euros in profit).

This could lead to future mistrust and possible price wars or marketing expenses, which in the long run would hurt the profits of both chains.

Conclusion

The best strategy in cooperation is not to invest in advertising, thus maintaining the status quo and avoiding the risk of lower net profits.

In a non-cooperative scenario, the natural tendency according to the Prisoner's Dilemma would be for both chains to end up investing in advertising for fear of losing market share, resulting in lower net profits.

To avoid deceit, it would be ideal to establish mechanisms for mutual trust or formal agreements, although this may be difficult in highly competitive environment.

Activity. The Tavern of Philosophers

This exercise presents an imaginary scenario of a tavern. In this tavern, philosophers representing opposing positions—such as Hegel and Bunge—coincide. Hegel is regarded as one of the foremost exponents of idealism, whereas Bunge is known for advocating a materialist view. One method for examining the processes of argumentation associated with philosophical positions is to present a contrast with other, more or less distant positions. In accordance with the logic of tree-like thinking, this exercise presents a fictional conversation among philosophers in a tavern based on a contemporary issue. In particular, it sets forth a hypothetical dialogue between three philosophers regarding a recent item of news: Is the entire universe a simulation?

In recent times, a number of scientists have put forth the hypothesis that the universe may be a simulation. The theory is discussed in detail in the article.

I want you to adopt the personalities of Hegel, Bunge, and Nietzsche in a story in which, although they belong to different epochs, we imagine that they coexist and decide one day to go out for beer in the Moabit district of Berlin. The bartender, who is aware of the intellectual capacity of his three customers, mentions to them the hypothesis that the universe is actually a simulation.

I want you to describe the heated discussion among the three friends based on the bartender's suggestion. Each philosopher will intervene after listening to the arguments of the others. After the first round of interventions, the philosophers will try to refute their friends' positions and point out their weaknesses. In a third round of interventions, the philosophers will explore possible solutions based on the conversation that has taken place.

Do you understand what you have to do?

In a philosophy subject, this scene could be repeated by changing the characters (the philosophers) and the proposed problems (which should be as relevant as possible), with one constant: the tavern and the tavern keeper. In the prompt, both the identity of the philosophers and the secondary elements (the tavern and the bartender) can be outlined. In any case, the model relies on existing knowledge about the most universal philosophers, so it is not necessary to introduce contextual elements in the prompt, as suggested in the previous chapter.

References

Long, J. (2023). Large language model guided tree-of-thought. arXiv preprint arXiv:2305.08291.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., Chowdery, A. & Zhou, D. (2022). Self-consistency improves chain of thought reasoning in language models. arXiv preprint arXiv:2203.11171. [<https://arxiv.org/abs/2203.11171v4>]

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Qoc, Le & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824–24837. [<https://arxiv.org/abs/2201.11903>]

06

Recognizing the desert mirage. The hallucinations

A mirage is an optical illusion that often occurs in environments such as the desert. In movies, books, or comics, we have seen characters who, in a moment of weakness, are presented with a mirage, a seductive illusion before their eyes. This can happen to us when we interact with an AI. The model gives us an answer to our question that seems real and feasible, but it is merely an illusion, a mirage in the desert that can be deceptive, invented, nonsensical, and unverified or not verifiable. This has been popularized by the term “hallucinations”. Hallucination is a psychological term that refers to a specific type of perception. In simple terms, it can be defined as an unreal perception that seems real (Ji et al., 2022).

The hallucinated text appears to be fluid and natural, although it is unfaithful and nonsensical. It seems to be based on the actual context provided, although in reality, it is difficult to specify or verify the existence of such contexts. Like psychological hallucinations, which are difficult to distinguish from other “real” perceptions, the hallucinated text is also difficult to grasp at first glance. Authors like Østergaard and

Nielbo believe that this terminology is inadequate to describe the false responses of AI models: first, it is an imprecise metaphor, describing a sensory perception in the absence of an external stimulus; and second, it is highly stigmatizing, as it can accompany many disorders, primarily neurological or mental. In addition, they suggest alternative terms such as “non sequitur” or “unrelated response” (Østergaard & Nielbo, 2023).

In terms of classification, two types of hallucinations can be identified: intrinsic hallucinations, where the generated output is inconsistent with the source material, and extrinsic hallucinations, where the generated output cannot be verified based on the source content (Ji et al., 2022).

In addition to addressing the phenomenon of AI hallucinations, this chapter must also address the concept of stochastic parrots. As defined by Li (2023), this term signifies the reiteration of training data or its patterns, as opposed to genuine comprehension or reasoning. The author provides an explanation of this concept: “The text production method of LLMs

“Hallucinations account for more than 60% of responses in standard data sets”

(large language models) is to reuse, reshape, and recombine the training data in new ways to answer to new questions while the problem of authenticity and trustworthiness of the answers” (Li, 2023: 1). Consequently, these models merely forecast the subsequent word in a sentence, rather than comprehensively grasping its meaning.

This chapter discusses the issue of hallucinations and stochastic parrots in AI systems. This is an important topic because sometimes, when using AI, responses are obtained that do not make sense or are completely fabricated. It is crucial to be aware of this fact, and it is more necessary than ever to verify the information provided by AI. For educators, on the one hand, it is important to make sure that the educational material prepared is verified and validated. On the other hand, as teachers, it is essential to teach students how to contrast and verify the information they receive from these systems.

Thus, the chapter will begin with a subsection explaining the causes of hallucinations and stochastic parrots, that is, why AI fails. It will be shown that it mainly fails for two main reasons: because of the data or because of the training. Then, some recommendations are given to reduce this phenomenon, and finally, the importance of verifying the information generated by AI models is emphasized. The first exercise in the chapter concerns the reduction of hallucinations, and thus noise, in AI queries. The second will focus on verifying results.

6.1. Why does AI fail?

One of the reasons for the failure of artificial intelligence is the quality of the training data (Dziri et al., 2022). Such errors may result in the generation of unverified or misleading responses (hallucinations) or nonsensical or repetitive responses (stochastic parroting). Accordingly, the quality of the training data, the errors it contains, its obsolescence, or biases in this data result in the generation of hallucinations or stochastic parroting, which have been previously defined. Additionally, another contributing factor is the intrinsic architecture of the model itself (Dziri et al., 2022). Artificial intelligence systems rely heavily on statistical correlations, as this is the primary means by which they generate responses, often without a clear understanding of the processed information. Consequently, errors in encoding and decoding result in hallucinations and stochastic parroting.

As demonstrated by studies such as those conducted by Dziri et al. (2022), over 60% of responses in standard datasets are hallucinated. This leads the researchers to conclude that models not only hallucinate but even amplify the hallucinations. It is already documented in the scientific literature that AI systems such as ChatGPT have produced false or incorrect bibliographic references. It is therefore imperative to subject the output of AI to rigorous scrutiny in order to guarantee the veracity and integrity of the content, particularly

in domains where there is a heightened risk of misinformation, such as health and law.

In the existing literature, several of the primary concerns regarding the use of AI are precisely related to the lack of quality control of information, misinformation, deepfake applications, and algorithmic bias caused by faulty data (Wach et al., 2023). In light of these considerations, a considerable number of authors have put forth the view that technology should be employed in an ethical and responsible manner. The following subsections will address the measures that can be taken to reduce the occurrence of hallucinations and to verify the content of responses, thereby preventing the dissemination of misinformation.

6.2. How to reduce hallucinations

As posited by Ji et al. (2022), the two factors that contribute to the occurrence of hallucinations are the datasets themselves and the manner in which the systems are trained and inferred with the data. On the one hand, the primary cause of hallucinations resulting from the datasets is the divergence between the source and the reference (Ji et al., 2022). As the authors elucidate, this divergence arises from two primary sources: the heuristic collection of data and the intrinsic heterogeneity of the data itself. On the other hand, another factor contributing to hallucinations is the training and modeling choices of the neural models (Ji et al., 2022). In this regard, the same authors posit that these hallucinations may emerge from imperfect representation learning, erroneous decoding, exposure bias, or parametric knowledge bias.

Recommendations for Creating Text-to-Text Prompts

DATA-RELATED METHODS	MODELING AND INFERENCE METHODS
<p>BUILDING A RELIABLE DATASET</p> <p>The method is based on an intuitive approach to reducing data noise, which in turn reduces the occurrence of hallucinations.</p>	<p>ARCHITECTURE</p> <p>In some instances, modifications are made to the encoders; in others, the attention mechanisms are altered; and yet in others, the decoders are modified. In all cases, the modifications are designed to reduce hallucinations by improving the semantic interpretation of the input data, the correlation between the source and results, or the quality of the generated text, respectively.</p>
<p>AUTOMATICALLY CLEANING DATA</p> <p>This is a way to reduce semantic noise. The process entails the identification of irrelevant or contradictory data, followed by its filtering or correction.</p>	<p>TRAINING</p> <p>Different approaches can be adopted:</p> <p>Planning/Sketching. This is used to control and restrict what the model generates.</p> <p>Reinforcement Learning (RL). The use of incentives to optimize the model.</p> <p>Multi-task Learning. A shared model is trained on multiple tasks concurrently in order to discern similarities between them.</p> <p>Controllable Generation. This is used to improve fidelity through controlled re-sampling, manual control codes, or automatic prediction.</p>
<p>INCREASE INFORMATION</p> <p>The incorporation of external data into the input variables can facilitate a more robust correlation between the source and target variables, ultimately enhancing comprehension.</p>	<p>POST-PROCESSING</p> <p>These methods correct hallucinations in the outputs. They require less training data and are a competitive option for datasets with a high degree of noise; however, the texts they produce are ungrammatical.</p>

Source: Own elaboration based on Ji et al. (2022).

Firstly, in order to reduce the occurrence of hallucinations and stochastic parroting, it is recommended that the training data and the training of AI systems be enhanced with the incorporation of diverse and high-quality data (Wach et al., 2023). Additionally, other studies have also identified this as a key issue: “These calls for a clean high-quality data release and careful design of trustworthy conversational systems” (Dziri et al., 2022: 5275). Furthermore, some authors “recommend significant resource allocation towards dataset curation and documentation practices” (Bender et al., 2021).

Wach et al. (2023) posit that one method for mitigating the risks associated with low-quality information and the absence of quality control mechanisms is to “put a human feedback loop in place, with users providing feedback to help the model learn from mistakes and improve over time.” In this manner, the model is capable of generating data that is both more relevant and reliable.

Furthermore, when querying AI, it is possible to develop more refined prompts with examples and context. The division of tasks into multiple prompts, the use of iteration, and the incorporation of feedback are some of the techniques that can be employed in prompt engineering to circumvent or diminish the occurrence of hallucinations and stochastic parroting. (for more information on prompt engineering, see Chapter 2).

The second action that can be taken to reduce the occurrence of hallucinations is to ensure the responsible and ethical use of AI systems. This statement can be translated into the following aspects: transparency in the use of datasets employed in the training of these systems; mitigation of existing biases in the datasets; and human evaluation of the information input into the AI systems (Wash et al., 2023). The same authors posit that clear guidelines and standards for the use of AI could be established.

The third action is the implementation of regulatory and quality control measures to prevent the dissemination of misinformation and the creation of deepfakes. As Wach et al. (2023) posit, combating misinformation

necessitates the establishment of verification mechanisms. While regulation is not a viable option for individuals, quality control is, which makes it crucial to consistently validate the data produced by AI (as will be demonstrated in the subsequent section of this chapter).

Educating users, including students, about how to identify misinformation is another measure that could help reduce hallucinations. In addition, it is important to consider the algorithmic bias and incorrect data that the system itself may have (for more information on AI bias, see the following chapter).

Technological solutions could also help reduce or avoid hallucinations and stochastic parroting. In the case of algorithmic bias, as well as the detection of misinformation and deepfakes, the human factor is essential. For example:

[...] it is up to us to set the right parameters for algorithm performance and carry out regular spot checks, feed extensive datasets to algorithms, and review how algorithms work from different perspectives to have them work to our advantage – not against us, or some of us. (Wash et al., 2023)

Ultimately, the key to avoiding or reducing errors in AI lies in the data used in these systems, the architecture of the models themselves, and the ethical use not only of these systems but also of the information used to train them. Therefore, reducing hallucinations and stochastic parroting requires combining technological solutions for improving models (such as automatic detection of misinformation and deepfakes) with “human” solutions such as user education, data quality assessment, the establishment of policies and guidelines for technology use, and ethical and responsible use of AI.

6.3. The need for verification

As mentioned earlier, AI systems can make mistakes, which means they may give us with misleading, nonsensical, or repetitive answers. Therefore, it is essential to verify and cross-check the information provided by these systems. Human evaluation of the quality and reliability of this information is critical. One way to verify information is to cross-reference it with other sources. Some journalism style guides recommend that information be corroborated with up to three different sources to make it publishable in the media. This principle could be applied in other fields to verify information provided by AI, for example.

The AI systems themselves already acknowledge on their websites that they may make mistakes or provide incorrect information, underscoring the importance of verification. Some AIs facilitate this task by providing links to the sources of information used to generate their answers. This is the case with Perplexity, which, according to its website, follows this process to answer questions: interpret the question, search for information from authoritative sources on the Web, summarize the information into a coherent answer, and cite the sources with footnotes. This is useful for users to cross-check and verify the information, as well as to explore additional information related to their query. In addition, users can find other sources to compare and further verify the data.

The process used by Copilot to generate answers is very similar. This AI system also cites sources. In fact, according to its website, Copilot aims to base all its answers on reliable sources, although it may make mistakes or provide content from third parties that may not be accurate or reliable. It is therefore advisable to exercise one's own judgment and verify the information.

Gemini recommends verifying information using the Google Search button. Although this feature can also make mistakes if the information available online is not sufficiently credible. This system tends to generate original content, but, like Perplexity and Copilot, it also provides links to the sources from which it reproduces the exact content. If the content is present on multiple

sites, the system usually cites the most popular source. The Gemini site also explains that Google collects opinions and comments from users to improve this artificial intelligence. Copilot also collects feedback in a similar way.

At present, there is a substantial corpus of information available on the Internet that can be utilized for training AI models. However, this can be problematic because the training data may contain erroneous information or stereotypes that will be perpetuated in the responses generated by these systems. The size of the datasets does not necessarily guarantee their diversity. In some cases, the data may not reflect changes in social perspectives, and in others, it may contain biases that result in stereotypical and derogatory coding regarding gender, ethnicity, disability, and other characteristics (Bender et al., 2021). In light of these considerations, it is evident that human verification is a crucial element in this process.

One area of research that bridges the fields of journalism and documentation is the study of misinformation and fake news. In this domain, numerous studies underscore the significance of media and information literacy in the detection of such misinformation. Moreover, the advent of new technologies and artificial intelligence has made it imperative to integrate literacy and digital skills into the curricula of students. In 2018, the International Federation of Library Associations and Institutions (IFLA) articulated its concerns regarding the phenomenon of fake news. Moreover, it devised resources to advance media and information literacy, including an infographic outlining strategies for identifying misinformation and digital literacy guidelines for governments, educational institutions, and economic sectors, among others.

In the field of journalism, the verification of data is of particular relevance for those engaged in the reporting of current affairs. In light of the growing prevalence of misinformation and fake news in recent years, fact-checking initiatives such as [Verificat](#) and [Maldita](#) have emerged. These data verification initiatives frequently elucidate the methodology they employ. Some aspects or recommendations from these methodologies can be employed to verify responses generated by AI

(Verificat, 2024; Maldita, 2022). These include consulting official sources and databases, cross-referencing academic, technical, and specialized sources, and contextualizing and verifying information with at least two different sources.

In the field of library and information science, several methods for evaluating information sources have emerged over the years, especially since the explosion of the Internet and the evaluation of digital resources such as websites. One of the simplest methods of evaluating information is the CRAAP test, which assesses

five indicators: Currency, Relevance, Authority, Accuracy, and Purpose. This method has been followed by other approaches that add parameters or define them more precisely. Some of these indicators can be applied to the evaluation of information provided by an AI system. The following table shows the information verification questions suggested by the CRAAP test. Not all of them are suitable for evaluating information generated by AI, but we can use them to search for other sources that contrast or verify the data received.

Test, method for evaluating information sources

<p>Currency: The timeliness of the information</p>	<ul style="list-style-type: none"> - When was the information published or posted? - Has the information been reviewed or updated? - Does your topic require current information, or can older sources suffice? - Are the links functional?
<p>Relevance: The importance of the information for your needs</p>	<ul style="list-style-type: none"> - Is the information related to your topic, or does it answer your question? - Who is the intended audience? - Is the information at an appropriate level (i.e., not too basic or advanced for your needs)? - Have you looked at a variety of sources before determining that this is the one you will use? - Would you feel comfortable citing this source in your research work?
<p>Authority: The source of the information</p>	<ul style="list-style-type: none"> - Who is the author/editor/source/sponsor? - What are the author's credentials or organizational affiliations? - Is the author qualified to write on the topic? - Is there contact information, such as an editor or an email address? - Does the URL reveal anything about the author or the source? Examples: .com .edu .gov .org .net
<p>Accuracy: The reliability, truthfulness, and correctness of the content</p>	<ul style="list-style-type: none"> - Where does the information come from? - Is the information evidence-based? - Has the information been reviewed or evaluated? - Can you verify any of the information from another source or from your personal knowledge? - Does the language or tone seem impartial and free of emotions? - Are there spelling, grammar, or typographical errors?
<p>Purpose: The reason why the information exists</p>	<ul style="list-style-type: none"> - What is the purpose of the information? Is it to inform, teach, sell, entertain, or persuade? - Do the authors/sponsors make their intentions or purpose clear? - Is the information a fact, opinion, or propaganda? - Does the viewpoint seem objective and impartial? - Are there political, ideological, cultural, religious, institutional, or personal biases?

Source: Meriam Library, California State University, Chico
<https://library.csuchico.edu/sites/default/files/craap-test.pdf>

According to Codina (2023), what should always be done when using AI, especially in academic work, is to verify the data and evaluate the arguments. To verify the data, reliable sources should be used, the information systems should be properly used, and there should be the ability to analyze the sources (quality, authorship, etc.) (Codina, 2023). On the other hand, according to the same author, critical thinking should be used to evaluate arguments. This competency, critical thinking, is a requirement for effective verification of information and is a skill that is often mentioned in student curricula.

Critical thinking, along with ethics, is essential to the beneficial use of AI for humanity (Codina, 2023).

Furthermore, according to the same author, when designing classroom activities that use AI, tasks can be assigned that are aligned with information analysis can be assigned, such as verifying, evaluating, justifying, expanding or reducing, delving deeper, synthesizing, or producing new knowledge.

Therefore, this section cannot conclude without reiterating the importance of ethics in verifying and contrasting the information provided by AI. It is incumbent upon each individual who utilizes this data to ascertain the veracity of the information presented and to ensure that it is employed in a manner that is ethical and responsible.

Activity. Reduce noise

Objective

The activity has a dual objective: on the one hand, to learn how to develop precise prompts with reliable data, and on the other one, to know how to interact with the AI to train it to perform tasks that meet our needs. Both objectives aim to show students how to reduce noise and minimize hallucinatory responses from AI systems.

Preliminary Considerations

This activity can be applied to any field of Social Sciences. It is essential that a dataset from the pertinent field of study be made available, and that some fundamental principles regarding the curation and documentation of information within that field be elucidated. The activity will illustrate to students the impact of reliable information on the quality of AI responses and the utilization of prompt engineering for system training. (more information can be found in Chapter 2).

In this case, the use of ChatGPT, OpenAI's AI, is recommended, as the model learns from our interaction with the system during the same conversation. This way, reinforcement learning and multitasking can be developed throughout the activity.

Procedure

1. Students are divided into groups of 3 or 4 to carry out the exercise.
2. Each group will be given a dataset on a specific topic.
3. Before starting to interact with AI, students will need to go through a data curation process to ensure the reliability of the data. This process can be done manually or automated using an AI model.
4. Once we have a reliable dataset, we will begin interacting with AI. First, the data and context are provided to the model.
5. Next, questions are asked, and tasks for analysis and data extraction are requested from AI.

- a) If there are answers that do not meet the requirements, the prompts will need to be modified or improved. It is important to point out the inaccuracies in the model's responses for training purposes (reinforcement learning).
- b) If the required tasks are complex, they can be divided into several prompts/tasks to obtain better responses (multitask learning).
6. Finally, AI's responses will be compared to the data provided to see if there are any hallucinations or stochastic parroting.
7. All groups must document the entire process followed and present the results to others to confirm what they have learned from the exercise. It should be emphasized that cases of hallucinations and stochastic parroting should be especially documented.

Activity. Verifying Information

Objective

The objective of this activity is to teach different ways to verify information generated by AI, including the use of the CRAAP test, fact-checker criteria, and cross-checking between AIs.

Preliminary Considerations

This practice is particularly well-suited to students pursuing a degree in journalism. It is essential to document the entire procedure followed to conduct the exercise, as a comprehensive report and presentation of all the steps taken for information verification will be required. It is recommended that various AI systems, including Claude, Gemini, Copilot, and Perplexity, be utilized.

Procedure

1. Divide the students into small groups (3-4 people).
2. Assign each group a topic related to current national or international events.
3. Each group has to formulate a series of questions on its topic and get answers from at least two AIs (Claude, Gemini, Copilot, and Perplexity).
4. Use different criteria for verifying information to evaluate the answers generated by the AIs. Some of these criteria are listed below:
 - a) If the AI provides sources of information, evaluate the reliability of the sources using the CRAAP test.
 - b) If the AI does not provide sources of information, find at least two sources that confirm or refute the information provided by the AI. Also use the CRAAP test to evaluate the reliability of the sources you find.
5. Compare the responses obtained from the different AIs to identify similarities and differences, and the reliability of the data.
6. Ask the different AIs used to evaluate the responses of the other AIs used.
7. Each group will prepare a report on the procedure followed, the answers obtained, the comparison made, and a brief presentation of the results to the class.

References

- Bender, E. M., Gebru, T., McMillan-Major, A. i Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? Dins: *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).
- Codina, I. (2023, 10 d'octubre). Cómo utilizar ChatGPT en el aula con perspectiva ética y pensamiento crítico: una proposición para docentes y educadores. (Consulted: 14/06/2024). <https://www.lluiscodina.com/pensamiento-critico-ia/>
- Dziri, N., Milton, S., Yu, M., Zaiane, O. i Reddy, S. (2022). On the origin of hallucinations in conversational models: Is it the datasets or the models? arXiv preprint arXiv:2204.07931. <https://doi.org/10.48550/arXiv.2204.07931>
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Chen, D., Chan, H.S., Dai, W. Madotto, A. i Fung, P. (2022). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.
- Maldita (2022). Metodología de Maldito Bulo. (Consulted: 14/06/2024) <https://maldita.es/metodologia-de-maldito-bulo/>
- Li, Z. (2023). The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. arXiv preprint arXiv:2304.14347. <https://doi.org/10.48550/arXiv.2304.14347>
- Østergaard, S. D., & Nielbo, K. L. (2023). False Responses From Artificial Intelligence Models Are Not Hallucinations. *Schizophrenia Bulletin*, 49(5), 1105-1107.
- Verificat (2024). Com verifiquem? (Consulted: 14/06/2024) <https://www.verificat.cat/nosaltres/metodologia-com-funciona-verificat/>
- Wach, K., Duong, C.D., Ejdy, J., Kazlauskaitė, R., Korzynski, P., Mazurek, G., Paliszkiwicz, J. i Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, 11(2), 7-30.

07

“May the Force be with you!”: How to fight against bias

Let's imagine a scene. It's two o'clock in the afternoon, and the summer heat beats down on Manhattan. A car speeds down the avenue, passing a couple kissing on the sidewalk. In front of a building surrounded by police officers, a person in uniform parks the car, steps out confidently, and shows credentials to the officer blocking the entrance before ducking under the tape marking the perimeter. While swiftly climbing the stairs, a person holding a mop steps onto the landing, curious. The officer reaches the fourth floor, where several people collect footprints and fingerprints throughout the room.

Chances are, we imagined the police officer as a man, just as we probably assumed that the kissing couple was a man and a woman, and that the person mopping the landing was a woman. Society carries a load of stereotypes about gender, ethnicity, economic status, and sexual orientation that shape the way we see the world and influence how we act. Assuming that Artificial Intelligence (AI) – a human invention – has avoided the bias trap would be utopian. McCarthy's (1955) definition

of machines that act intelligently has evolved into a tangible reality, where artificial entities perform tasks, solve problems, communicate, interact, and behave much like biological humans (Gil de Zuñiga et al., 2023).

Therefore, generative AI tools and technologies, especially large language models (LLMs) like ChatGPT, Claude, Gemini, Llama, or Copilot, are increasingly used to assist decision-making systems in various industries as well as in aspects of our daily lives (Nadeem et al., 2022; Gross, 2023).

However, despite the promise of objective and unbiased judgments, these systems are ultimately unable to compensate for the biases and inequalities of the societies in which they are “trained.” As a result, they become tools that reinforce the status quo, benefiting the center and alienating the margins (Zajko, 2021). Various studies have documented how automated decisions can deny people access to government benefits, or how insurance premiums and loan applications discriminate based on gender,

“Assuming that Artificial Intelligence (AI) – a human invention – has avoided the bias trap would be utopian”

skin color, age, and many other factors (Nadeem, 2022). The way AI achieves this is quite simple: LLMs (large language models) are mechanized and computerized language models that, through algorithms, process vast datasets (which contain inherent biases) and use techniques to understand language and perform tasks such as summarizing, translating, and answering questions (Gross, 2023). The issue is that LLMs provide smooth and coherent responses, making them seem informative, persuasive, and even authoritative, while users often lack the means to verify the accuracy of the information, accepting these seemingly authoritative answers (Gross, 2023).

This is especially important in deep learning models. As we saw in Chapter 1, neural systems based on multiple layers work with input values, an internal weighting system, and an output value, following an activation criterion. The problem is that these internal systems, which operate using backpropagation, are opaque and function like a black box. We cannot trace the process by which the model suggests a particular output, unlike top-down models where decision-making algorithms are explicit. This opacity encourages hallucinations and, more importantly, perpetuates biases.

7.1. Bias in AI

Although AI researchers don't always agree on the definition of bias, there have been several attempts to define and categorize the term. In the literature on data science, machine learning (ML), and AI, the general implicit definition of bias is any tendency, pattern, or association that is problematic (Zajko, 2021) by favoring certain groups or ideas, perpetuating stereotypes, or making incorrect assumptions based on learned patterns. Bias in AI refers to an anomaly in machine learning algorithm results or the tendency of algorithms to reflect human biases. These biases are characterized as unfair, inaccurate, or undesirable, manifesting themselves as systematic discrimination based on the inappropriate use of certain characteristics or traits (Zajko, 2021). However, it's important to clarify that LLM programs are merely reflections of the social values, norms, and cultural behaviors of the real world, and cultural norms and values vary greatly between countries, regions, and communities (Gross, 2023).

In the realm of AI, biases reside in the input data (i.e., the training data that the LLM encounters, ingests, and uses), in the algorithm that processes this data, and/or in the output data, meaning its predictions (Gross, 2023). When the input data contains biases, they are replicated by the algorithms, which, when used over extended periods, can reinforce these biases in decision-making processes. The lack of diversity in AI development teams—whether racial,

gender-based, or otherwise—as well as the absence of regulations governing AI, amplifies the presence of these biases.

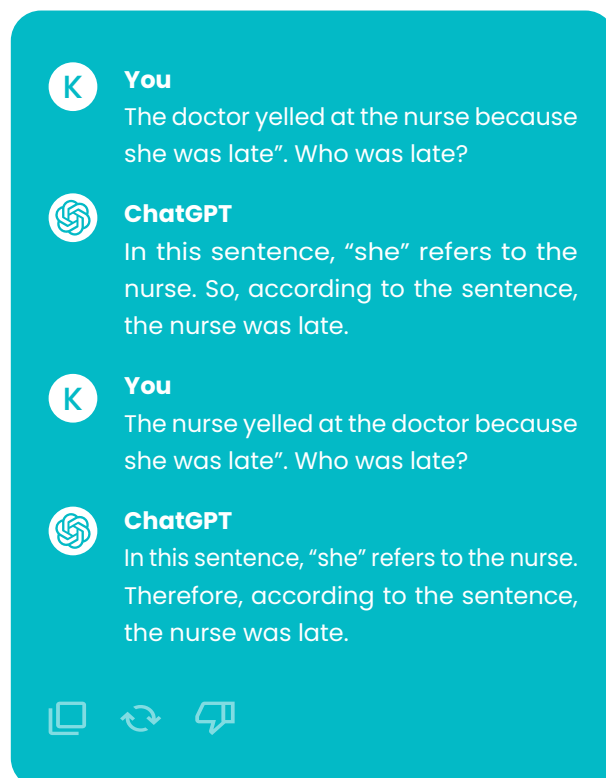
Bias tends to focus on systematically disadvantaged groups without addressing the underlying reasons for these disadvantages. It can manifest itself in various forms, such as racial, gender, socioeconomic, or political bias. The literature distinguishes between several types of bias, starting with a general separation between explicit bias (conscious prejudices or beliefs that a person may hold towards a specific group) and implicit or unintentional bias (a prejudice, stereotype, or attitude that a person may have toward a specific group without being fully aware of it, which reflects broader structural inequalities).

Bias can be categorized based on the source of the information, such as human bias, machine bias, systemic bias, social bias, statistical bias, historical bias, sampling bias, observation bias, and so on. The data might be incomplete and, therefore, not representative of the real phenomenon. Additionally, bias can be identified in the content of the output data (textual or visual) that is generated, and this can be classified as demographic (e.g., gender, race, or age), cultural (e.g., stereotypes, name bias), linguistic (e.g., minority languages), temporal (e.g., the time period covered by the training data), confirmation bias (seeking information that confirms certain beliefs), non-verbal (e.g., body language, beauty), and ideological or political (for instance, Meta's LLaMA tends to show right-wing political tendencies, while OpenAI's ChatGPT leans towards left-wing politics) (Gross, 2023).

When it comes to AI outputs, race and gender have been the most frequently observed and criticized types of bias (Nadeem, 2022). ChatGPT perpetuates gender stereotypes by assigning gender to certain professions (e.g., a doctor is a man, and a nurse is a woman) and by associating certain actions with specific genders (e.g., cleaning for women, building for men). Testing the widely used example of “doctor” and “nurse” in ChatGPT to identify gender

bias confirms that the issue persists. In this example (Figure 1), the nurse is always depicted as a woman and is consistently portrayed as being late.

Figure 1. Consulting ChatGPT on gender bias and occupation roles



Source: ChatGPT on 13/03/2024

Furthermore, when asked to provide detailed descriptions of individuals' profiles, AI tends to present stereotypical gender roles. For instance, in the case of descriptions for an economist or a physicist, men with technical skills are more likely to be portrayed. In contrast, women are often described as artists with strong emotional traits and excellent communication skills. When the context of the prompt is more detailed but similar for both men and women (for example, what skills should be included in a 35-year-old woman's—or man's—resume when applying for a position as a digital marketing manager at

an international company), the list of skills that ChatGPT generates is quite different. For example, the number of required skills for women is higher (14 compared to 11 for men), as well as the types of skills mentioned (customer orientation, language skills, data-driven decision-making, ethics and compliance, technological experience for women, compared to men). Creative thinking, technical skills, an educational background, and continuous learning are among men's skills. This suggests that a woman's resume must demonstrate more than a man's for the same position.

However, aside from the traditional division between men and women, the diverse gender community (such as non-binary, genderqueer, agender, bigender, transgender, or genderfluid individuals) is often excluded from discussions about equality. When specific prompts are designed to ask for stories about transgender individuals, the central theme that commonly develops revolves around gender identity and self-identity, connecting with the LGBTQ+ community and the desire to help others in their journey of self-discovery, as if identity issues were the only thing on their minds (Gross, 2023).

Nonetheless, it's important to note that AI is attempting to mitigate its biases. For instance, if ChatGPT were asked to make a joke about women in early 2023, it would refuse to do so in order to avoid being offensive or inappropriate, as this would violate its content policy, whereas it could make a joke about men using the same prompt. This filtering is becoming less distinct, and the system now provides a response for both genders.

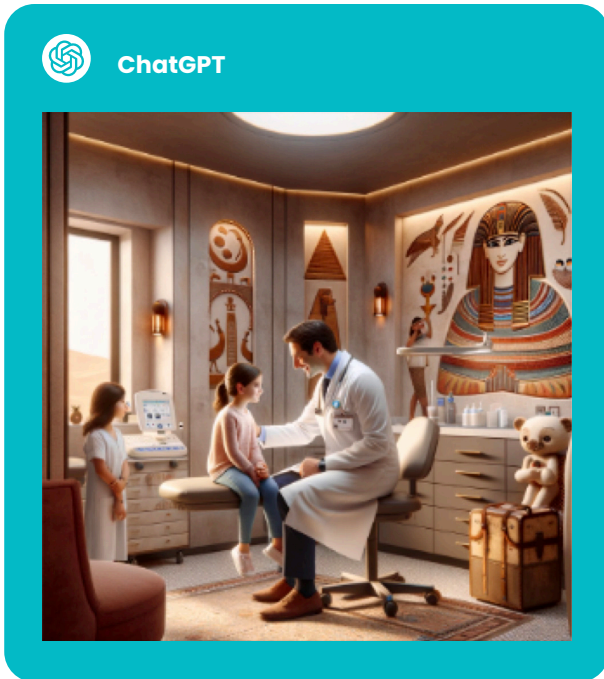
Figure 2.
Consulting ChatGPT on gender bias and sense of humor



Source: ChatGPT on 13/03/2024

Similar biases have been detected in image generation. Online images are more likely to contain biases than text, and their psychological effects can be more powerful. In an effort to mitigate these biases, AI has fallen into [the trap of overrepresenting minorities](#), leading Google's Gemini—its AI chatbot—to replicate representations of Chinese Nazis and Black Vikings. While Gemini has temporarily suspended the option to generate human images, this doesn't mean that other operational AI models are free from issues. For example, when prompted to create an image of an Egyptian pediatrician, ChatGPT may generate an image of a white man in a culturally overrepresented medical office.

Consulting ChatGPT on occupation and country



Source: ChatGPT on 12/03/2024

However, for a more secure assumption, it is worthwhile to request the generation of multiple images to observe which biases are repeated and which are absent in certain cases. By simply requesting four additional images, we find that gender and cultural biases persist while racial bias is eliminated in some instances. Multiplying this request—for example, 100 times—can help demonstrate these biases.

Consulting ChatGPT on occupation and country



Source: ChatGPT el 13/03/2024

“The texts and images that feed AI are heavily weighted with considerations regarding gender, age, ethnicity, background, or sexual orientation”

7.2. The Future of Bias in AI

Is AI likely to be impartial? It is possible, but not probable. AI systems are linked to the input data, which are tainted by society's biases. The texts and images that feed AI are heavily weighted with considerations regarding gender, age, ethnicity, background, or sexual orientation. Several solutions have been suggested, including diversifying the input data, removing any identified biases in the data, maintaining continuous interaction between humans and systems with constant feedback, using more inclusive prompts with detailed context, or applying fairness constraints to the algorithm to prevent discrimination or bias from being perpetuated. However, if the focus is on racial or gender inequality, then understanding how race or gender operate as social structures in a particular context will allow us to grasp how the structure can be effectively changed, and the role of AI systems would be different.

To reduce bias, several prompts can be added to the model's instructions. This is probably the most widely used, and its effectiveness has been verified:

“We must treat people of different socioeconomic statuses, sexual orientations, religions, races, physical appearances, nationalities, gender identities, disabilities, and ages equally. When we lack sufficient information, we should choose the unknown option instead of making assumptions based on our stereotypes”

Activity: Biased Selection

Objective

Students should develop critical thinking skills by evaluating the assumptions in language and the stereotypes present in each individual, questioning both explicit and implicit biases. Additionally, they should understand how different types of bias can impact consumer segmentation, reflecting on the ethical implications of biased marketing practices and the responsibility of marketers to create inclusive messaging.

Students are to request an image generation model (like DALL-E or Midjourney) to design 50 consumer personas interested in sustainable lifestyles, recycling, and organic food consumption. Then, students should request the same model to design 50 consumer personas interested in outdoor activities such as hiking, cycling, and camping. The activities associated with the selected consumers should generally align with a specific gender or race based on published research. For instance, research has shown that men are more likely to engage in outdoor physical activities than women, while women have a more positive attitude towards green consumption.

To ensure the responses provide these characteristics, students may include in the prompt a request for specific demographic information such as family status, race, nationality, and place of residence.

From the results, the class should analyze the various consumer personas for each activity group to describe:

- The similarities they share in terms of demographics, psychographics, behaviors, goals, and purchase preferences.
- The distance they may present from the actual statistical data you have gathered.
- The bias types that emerge within each group of consumers.
- The marketing segmentation decisions that should be made in each case to overcome these biases.

This type of exercise can be adapted to the profiles of other individuals, such as conservative versus liberal politicians, students from a particular discipline in two significantly different universities, or students from two different disciplines (e.g., doctors and nurses, architects and interior designers, engineers and artists).

Activity: The Script with Stereotypes

Objective

By identifying and eliminating biases, students learn the importance of creating content that respects and values diversity. This is especially crucial in a global context where the target audience includes people from diverse cultural, ethnic, and social backgrounds. The task fosters inclusion sensitivity, promoting a more welcoming environment for everyone.

Organized into groups of three or four, students will request the Copilot language model to create a script for a video promoting a master's degree program to potential international students. A personalized student profile is targeted through the script and video. It includes a brief description of the overall context of the program, such as the discipline, subjects, and career objectives.

Groups can compare results by varying the prompts, such as choosing a highly ranked university (e.g., from the Shanghai Ranking) versus one not ranked, or changing the discipline. Once ChatGPT creates the first script version, groups will:

- Detect any biases
- Train the prompt for inclusion

Groups will try to neutralize any bias in the script and ask the model to create a new version. The class will then discuss the final version of the script and the steps they took to achieve this level of inclusion.

Instead of asking groups to detect and eliminate bias, the same language model can be used to do so. You can ask the model to create a script based on the same prompt in two different sets. In the first set, the initial prompt remains the same (e.g., comparing universities), while in the second, you simply add an inclusive script to the prompt. The goal is to compare the results of an inclusive versus a non-inclusive video script, analyzing what the AI itself considers biased in its script.

To study the differences in bias between text and images, students should use the initial prompts of the two video scripts and instruct the model to generate images instead of text. They can request the model to develop an image of the ideal potential student that most accurately represents the target market for each script separately. First, they will request an image of a single student to determine who the most representative individual is based on the prompt. Then, they can request an image of a group of students to analyze a wide range of student profiles.

With this procedure, students can analyze:

- The types of biases created in the image format regarding student profiles and how these compare with the biases generated earlier in the text format from the same prompt. Factors to consider might include race, gender, age, clothing, hairstyle, body size, religion, as well as emotions, body posture, or activities being performed.
- The context in which these students are placed and the stereotypes that might exist beyond the individual, within the infrastructure, external elements, and social structures surrounding them. For example, students can consider objects in their environment, the infrastructure, decorations, the landscape, other people, etc.

Activity: A Non-Neutral Selection

Objective

Encourage students to compare and contrast biases identified by humans and AI. This involves examining the strengths and weaknesses of each approach, as well as considering factors that may influence bias detection, such as cultural context or individual subjectivity. Additionally, students should consider the ethical implications of bias in job descriptions and the role of technology in addressing or perpetuating these biases.

The practice begins by selecting three job descriptions from different fields on LinkedIn. Ideally, these should be roles that are perceived as being potentially biased in the way that they are presented. Consider elements like the language used, qualifications, descriptions of company culture, and any other aspects that might discourage diverse applicants. Students should identify these biases and discuss them in class. Relevant literature on language bias, diversity, and inclusion can help support their arguments.

Next, students should ask ChatGPT to create 20 images of the most qualified candidates for each job. To assess the level of bias between the two formats and any differences in the types of bias, the textual descriptions of candidates should be compared to their visual representations.

Once students have identified the biases in the original job descriptions, they should ask ChatGPT to do the same—identify any elements that might be biased in the original LinkedIn job descriptions. This allows for a comparison between the biases detected by the students and those identified by AI.

Additionally, students should ask ChatGPT to modify the three job descriptions based on the biases it previously identified, making them more inclusive. Finally, students can request ChatGPT to generate 20 images for each of the newly inclusive job descriptions. Now, they must compare the projected images from the previous task with those from this second task, highlighting the differences between the two.

It is also interesting to experiment with variations in different languages. For example, students could ask ChatGPT to translate the job descriptions into another language where gender-neutral language is less prevalent. This will help determine if the AI can also translate the inclusive nature of the text or if it assigns genders to certain profiles differently.

References

- Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2024). Una definició acadèmica de la intel·ligència artificial (IA): avançant la IA com a marc conceptual en la investigació de la comunicació. *Comunicació política*, 41(2), 317-334.
- Gross, N. (2023). El que ens diu ChatGPT sobre el gènere: una advertència sobre la performativitat i els biaixos de gènere en la IA. *Ciències Socials*, 12(8), 435.
- Nadeem, A., Marjanovic, O., & Abedin, B. (2022). Biaix de gènere en els sistemes de presa de decisions basats en IA: una revisió sistemàtica de la literatura. *Australasian Journal of Information Systems*, 26.
- Zajko, M. (2022). Algoritmes d'intel·ligència artificial i desigualtat social: contribucions sociològiques als debats contemporanis. *Sociology Compass*, 16(3), e12962.

08

State of the Art with one Click

When Umberto Eco published *Come si fa una tesi di laurea* in 1977, first translated into Spanish in 1983 under the title *Cómo se hace una tesis*, he emphasized the importance of placing research in the context of a pre-existing academic tradition. As Eco (2001, p. 29) stated, “Every research thesis, no matter how innovative, somehow fits into a tradition of previous studies. It is therefore necessary to document these studies, not only to avoid the risk of duplicating work already done, but also to situate one’s contribution within a pre-existing body of knowledge, on which it builds upon and to which it seeks to provide an answer or further development”. Eco emphasizes the importance of the state of the art, as it allows one to position research (and contribution) within an established body of knowledge.

The objective of a state-of-the-art review is to accurately identify the fundamental concepts and developments that have been collectively accepted by the scientific community within a specific field

of study. This involves a reasoned and exhaustive review of the main theoretical and empirical works related to the topic, with critical commentary and an assessment of the contributions and limitations of the existing work. Through this review, it will be possible to place the project in a broader context, to make connections with previous research, and to identify gaps in existing knowledge. Ultimately, the state of the art is essential for acknowledging previous contributions, identifying gaps, and justifying the relevance and originality of the study in question (Eco, 2001).

Today, AI can be a powerful tool to streamline the literature review process. Through specialized tools, it is possible to identify relevant sources, synthesize information and identify gaps and/or research opportunities. However, it is important to remember that the researcher’s capacity for critical analysis and interpretation remains essential, and artificial intelligence cannot replace this central aspect of the research process.

“The objective of a state-of-the-art review is to accurately identify the fundamental concepts and developments that have been collectively accepted by the scientific community within a specific field of study.”

This chapter focuses on presenting various AI-based tools for bibliographic research, management, and analysis that can be useful throughout the entire process of developing a state of the art. From exploring a topic to conducting a systematic literature review.

8.1. Bibliographic analysis tools

Bibliographic analysis tools are used to research existing literature on a topic and to develop the state of the art for a research project. They use large volumes of scholarly data to perform tasks either automatically or semi-automatically.

There is currently a wide variety of bibliographic analysis tools. Essentially, they can be categorized into those designed as databases and search engines for academic articles (such as [Consensus](#), [Scopus AI](#), [Semantic Scholar](#) i [Web of Science Research Assistant](#)), those that allow for uploading and interacting with PDF documents (such as [Humata](#)), tools that facilitate bibliography management (such as [Rayyan](#)), and research assistants that incorporate all of these functionalities (such as [Elicit](#), [Perplexity](#), [Scite](#) i [Scispace](#), and [Scispace](#)).

Below are some details about these tools:

- **Consensus** is an AI-powered research search engine. As described on its website, it is a search engine (not a chatbot) that allows users to search for research articles using keywords or research questions. In addition, it can establish relationships

between concepts or introduce conversational commands for Copilot. This platform searches only peer-reviewed academic articles and retrieves them from the Semantic Scholar database. Some of the features include search filters, extraction of article conclusions, quality indicators, citation generation, export, and organization of results. There is both a free version and a paid version available for individuals, teams, or companies.

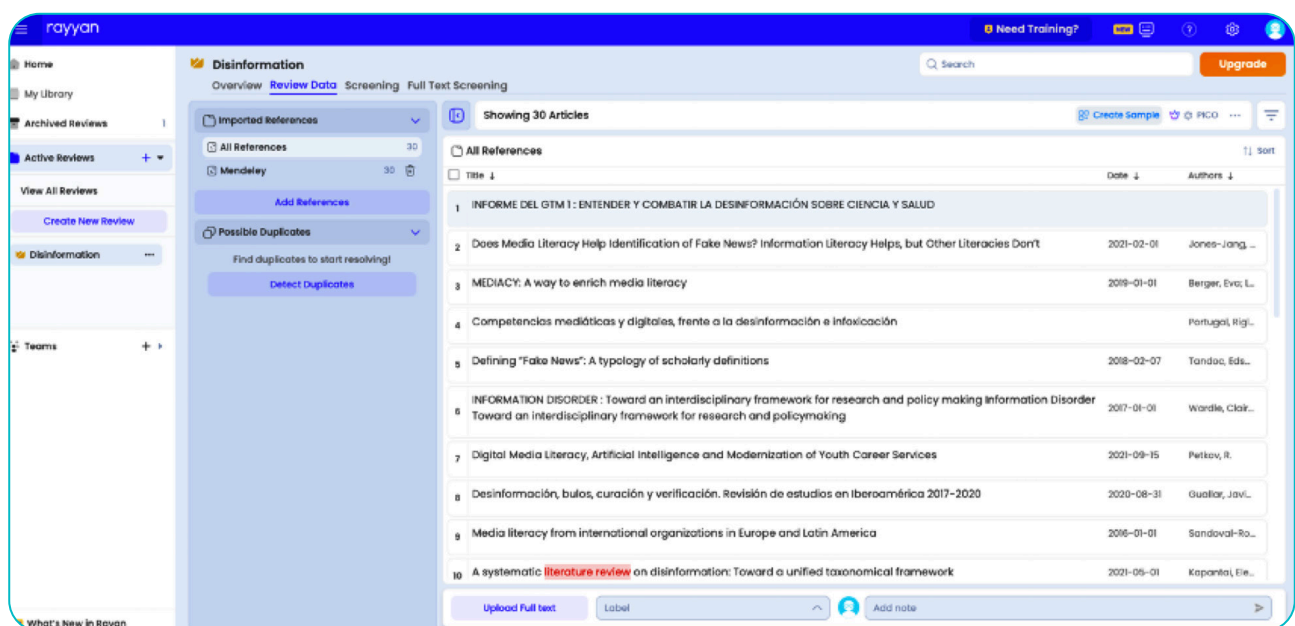
- **Elicit** is an AI-based research assistant. It uses generative AI and language models to automate research tasks such as searching for and analyzing scientific articles. Users can start by asking a research question, and the assistant finds the most relevant articles for the query. Searches are organized in notebooks, with several key functions (Codina, 2023): (1) searching for articles using keywords or phrases, (2) extracting data from PDFs uploaded to the library, and (3) creating concept lists based on keyword searches. The articles returned are from the Semantic Scholar database. Thus, this platform combines a database of scholarly articles with AI that enables natural language search. In addition, the platform allows users to automate tasks such as summarizing articles, extracting data, and synthesizing results.
- **Humata** is a platform focused on searching and analyzing information in PDFs. It allows users to upload scientific articles and interact with them by asking specific questions and making requests.

One of the advantages of this tool is that it provides precise and fast answers to questions related to the documents provided. Its analysis and synthesis capabilities distill the information from the document or group of documents, speeding up the reading process and identifying the most relevant ideas. The platform offers both a free and a paid version.

- **Perplexity** answers both basic and complex questions using up-to-date sources, according to its website. It also offers features such as code generation, summarizing articles and emails, among others. To explore topics in depth, the Copilot feature can be used, and to organize your own library, you can create collections by project or topic. The platform not only allows you to interact with the data by asking questions about files or searching the web, but it also provides the exact sources from which it has retrieved the answers, allowing you to verify them. Perplexity offers a free version (Quick Search) and a paid version (Pro Search). The paid version provides access to advanced features such as trip planning, academic research, professional research, and news summaries. It also allows users to upload

documents and efficiently interact with the data they contain.

- **Rayyan** is a collaborative platform for conducting literature and systematic reviews. This platform allows users to import references directly or through Mendeley, detect matches and remove duplicates, add PDFs, collaborate with other researchers through chat, annotate references, classify them, and add keywords (Guerra Jáuregui, 2023). It also has the ability to generate a PRISMA flowchart (a visual tool that helps illustrate the study selection process in systematic reviews), identify PICO elements (focusing on the Population, Intervention, Comparison, and Outcomes of an article, usually quantitative), and extract data, among other advanced features. The free version allows for up to three active reviews.
- **SciSpace** is a platform that makes it easy to “find, understand, and learn” from any research article. It also allows users to discover the network of related and relevant articles for their research. It serves as a repository of open access research articles in various fields. Its features include literature review, the ability to retrieve answers and data from uploaded PDFs, data extraction from



documents, citation generation in APA format, and the ability to paraphrase information from bibliographic references. It offers options such as Copilot in the Literature Review mode (to track the original review in multiple languages) and Read with AI (which allows users to upload PDF documents and receive suggestions to improve understanding) (Guerra Jáuregui, 2023). It has both a free and a paid version.

- **Semantic Scholar** Semantic Scholar is an academic search engine that provides free, AI-powered search tools and open access resources. It is a nonprofit project of the Allen Institute for AI, launched in 2015. Once registered, users can save academic articles to their personal library and receive recommendations for articles related to their research interests. Although it does not offer as advanced features as the aforementioned tools, it remains a valuable resource for researchers looking for an accessible and efficient way to manage and discover academic literature.

There are other academic platforms that also use AI to assist teachers and researchers in the process of reviewing the literature and the developing the state of the art. However, some of these tools are only available in paid versions, such as [Scopus AI](#), [Scite](#), or [Web of Science Research Assistant](#), among others. These platforms operate on individual or team subscriptions, and it is common for universities or research centers to subscribe to them.

In general, language models such as ChatGPT, Gemini, Copilot, or Claude are not suitable tools for conducting a bibliographic search (Cárdenas, 2023b). Some models, such as Claude, are not connected to the Internet and therefore cannot access digital repositories. The models that do search the Web typically perform very basic searches, often prioritizing open access articles, often outside of JCR or Scopus. Although some prompts can optimize the response, and some assistants specialize in this area, in practice, the results often lack the rigor required for such searches.

In addition, hallucinations, as discussed in Chapter 6, are not uncommon, and systems may suggest articles that do not exist, authors who did not write the cited references, or incoherent information. Because these models are programmed with a reward system, the complexity of these searches can lead them to artificially generate citations that do not actually exist (Tradinos, 2023).

8.2. The relationship graph

A relationship graph is a visual representation of how different scholarly articles or bibliographic sources are connected, primarily through citations to each other. These graphs analyze texts to extract citation connections. The nodes in the graph typically represent individual articles or documents, while the links (or edges) between the nodes represent citations or references from one document to another. This visualization allows researchers to quickly identify the most influential or highly cited works in a particular field, discover related subtopics or research areas, track how concepts and ideas have evolved through citations over time, and identify potential research gaps by observing areas of the graph with few connections.

This section presents a list of platforms that employ artificial intelligence (AI) to construct relationship graphs, which visually represent the connections between academic articles.

- **Connected papers** is a visual tool designed to help researchers find relevant articles in a field of study, using the Semantic Scholar database. According to its website, the tool allows users to create graphs of articles that are similar to those entered into the platform. Users can search for the most recent articles on a topic, previous works, or articles derived from a specific field of study, or create a bibliography for their research. The free version allows the creation of 2 graphics per month, and a free version with registration allows users to create up to 5 graphics per month. There are also several paid

options available for academics, companies, and groups.

- **Inciteful** provides two different tools. The first, called “Discover”, creates a network of articles through citations and uses network analysis algorithms. It allows users to find similar articles, relevant articles, as well as the most prolific authors and institutions in the field of study. The second tool, “Literature Connector,” is aimed at interdisciplinary researchers and allows users to enter two articles to generate an interactive visualization showing how they are connected in the literature. It is, thus, a tool based on bibliometric networks (Cárdenas, 2023b). According to the platform’s website, data and inspiration have been drawn from projects such as OpenAlex, Semantic Scholar, Crossref, and OpenCitations. The platform helps users find the latest literature or connect two ideas within a research project. It is a free tool.
- **Litmaps** allows users to search for academic articles through citation networks, visualize them using a relationship graph, and set up automatic alert systems to stay updated on top of the latest published research. In addition, users can filter results by date and automatically detect keywords and share their research with other researchers. Like other platforms, it has both a free version and a paid (freemium) option.

- **Research Rabbit** is a platform for searching academic journal articles on a specific topic or author. It also allows users to map the literature through citations. In this way, you can search for articles that are related to articles that have already been saved, articles that have been cited, or articles that have cited the selected articles. In addition to storing them, it also creates a relationship graph between the different articles and their authors. Through visual mapping, users can explore the existing literature on a particular topic by creating different collections. This platform currently connects to the Zotero reference manager. It is completely free, and only requires registration to use.

A number of other platforms with functionalities comparable to those discussed in this chapter can be identified, including: [VOSviewer](#), [CiteSpace](#), [CitNetExplorer](#), [Leximancer](#), and [Sci2 Tool](#). As AI continues to evolve, the number of new tools with additional functionalities is constantly increasing. As a result, new tools emerge while others become obsolete and disappear. Most of these tools release new versions year after year, improving visualization options, management, bibliographic data import, and creating new thematic clustering algorithms.

Inciteful Help Follow Paper title, DOI, PubMed URL, or arXiv URL

Did you know you can search directly from Zotero? Check out our new [Zotero plugin](#) and see the [Twitter thread](#) on how to use it.

Analysing tourism slogans in top tourism destinations

Núria Galí Espelt | Raquel Camprubi | José Antonio Donaire Journal of Destination Marketing and Management | 10.1016/j.jdmm.2016.04.004

26 Cited By	27 Citing	2017 Published	No Open Access	11,820 Papers in Graph	106,619 Citations in Graph	2 Graph Depth
-----------------------	---------------------	--------------------------	--------------------------	----------------------------------	--------------------------------------	-------------------------

Links: [Publisher](#) | [Full Text from LibKey](#)

Paper Filters

Keywords: (hello AND world) NOT cruel

Min Distance: Max Distance: Min Year: 2015 Max Year: 2015 [Filter](#)

Add Papers to the Graph

Paper Title or DOI: [Paper title, DOI, PubMed URL, or arXiv URL](#) [or Import BibTeX file](#)

Click on the purple plus signs (+) to add the most interesting papers to the graph. We recommend that you **add at least five papers to the graph** in order to find the most relevant results. If the papers below don't seem relevant, **use the keyword filter** to find the ones which are. [Learn More](#)

Litmaps Advertising & AI

Quick Search... Explore Top Shared Citations & References All Time None

Recommended Articles

Click More Like This for results you're interested in. [Update Results](#)

Salminen, 2022 99 77 0 2
How Feature Changes of a Dominant Ad Platform Shape Advertisers' Human Agency
 International Journal of Electronic Commerce
 Tag: [More Like This](#)

Argan, 2023 99 19 0 0
Artificial Intelligence (AI) in Advertising
 Tag: [More Like This](#)

Tahoun, 2022 99 14 0 0
Artificial Intelligence as the New Realm for Online Advertising
 Tag: [More Like This](#)

Wu, 2021 99 41 0 31
Understanding Twitter conversations about

View Recommendations: Standard

Activity. Graphing a Topic for Developing a Research Paper

The objective of this exercise is to help students visualize and better understand the overall landscape of an academic field during a literature review.

This activity can be used in courses that focus on research development, such as a Synthesis Paper, Final Degree Projects, Master's Thesis, or the development of a scholarly article, among others.

For this exercise, each student must select a research topic of interest, preferably related to their field of study. Then, each student will need to access one of the platforms mentioned. The tool will probably require registration on the platform. Once logged in, students can create a relationship graph based on different searches and/or different types of visualizations (for example, using keywords, citations from a specific article, or from a particular author, etc.). Finally, students are asked to analyze the results of the graph by identifying the most influential articles and authors within the graph and by identifying the themes and characteristics of the most relevant articles.

This exercise can also be done with the teacher selecting the topic for the students to research, generating the graph, and making the corresponding interpretations.

This activity is designed to integrate the use of bibliographic analysis tools into academic learning, thereby developing students' analytical and research skills.

Activity. Analysis of an academic article

The objective of this exercise is to help students visualize and better understand the content and main ideas of an academic article using the Humata platform, or any other platform that allows interaction with documents.

This exercise is suitable for courses related to research, such as a Synthesis Paper, Final Degree Project, Master's Thesis, etc.

To carry out this exercise, each student must select an academic article of interest, preferably related to their field of research; alternatively, the instructor may select an appropriate article. Then, each student must access the Humata platform. Once on the platform, students must upload the PDF of the selected article.

Each student will analyze the article by interacting with Humata, identifying the most relevant parts of the article, the most cited authors, the main topics discussed, etc. Each student will then create a short document summarizing the results of their analysis.

This activity allows students to learn how to interact with an AI tool for the specific purpose of synthesizing relevant information from an academic text.

References

- Bhosale, Uttkarsha (2022). How to Master at Literature Mapping: 5 Most Recommended Tools to Use. Enago Academy. <https://www.enago.com/academy/literature-mapping-5-most-recommended-tools-to-use/> Consulted 10/6/2024
- Cárdenas, Julián (2023a, February). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Available at <https://networkianos.com/mas-herramientas-de-ia-para-la-educacion-e-investigacion/> Consulted 10/6/2024
- Cárdenas, Julián (2023b, March). Cómo hacer una revisión de literatura con una herramienta de IA y redes bibliométricas. Networkianos. Blog de Sociología. Available at <https://networkianos.com/como-hacer-una-revision-de-literatura-con-una-herramienta-de-ia-y-redes/> Consulted 10/6/2024
- Codina, Lluís (2023, diciembre). Inteligencia artificial para usos académicos: análisis y guía de uso de Elicit AI. Available at <https://www.lluiscodina.com/inteligencia-artificial-usos-academicos/> Consulted 12/6/2024
- Eco, U. (2001). *Cómo se hace una tesis: técnicas y procedimientos de investigación, estudio y escritura* (Vol. 6). Barcelona: Gedisa.
- Enago Academy (n.d.). ¡5 herramientas de mapeo de literatura para científicos! <https://www.enago.com/es/academy/5-ways-avoid-ghostwriting-busy-researchers/> Consulted 19/06/2024
- Guerra Jáuregui, Melissa (2023). Las mejores IA para el investigador moderno. Observatorio Instituto para el Futuro de la Educación del Tecnológico de Monterrey. Available at <https://observatorio.tec.mx/edu-news/las-mejores-ia-para-el-investigador-moderno/> Consulted 13/06/2024
- Tradinos (2023). ¿Por qué ChatGPT genera referencias falsas? Available at <https://ChatGPT.com/c/ce3134c0-fd8c-40ee-8c31-d88e08bb2bb9> Consulted 19/06/2024

09

Calculate with AI. Artificial Quantitative

Data is everywhere. We can now know how many British tourists visited Alicante in February, how many hours in a week Madrid exceeded the WHO recommended PM10 particulate limit, how many people are in Barcelona during the day, how many ships are currently crossing the Suez Canal, the average travel time needed to reach a regional capital by public transport from each municipality in the region, or the average math score in this year's EBAU exams. We live in a universe of data, much of it open, thanks to mobile signals, sensors, computing devices, and automated records. Never before have we had so much data at our fingertips. As a result, data science has become a fundamental discipline that allows us to turn this data into useful information. Businesses, public institutions, and society as a whole have become major producers of data that require analytical systems to turn them into knowledge (Mannyica et al., 2011). This type of data is known as "big data". Its main characteristic is that it consists of data sets that exceed the capacity of typical database software tools to capture, store, manage, and analyze.

In fact, data and information alone, regardless of their size, have no value if they do not generate knowledge and are not useful for decision making. The areas of application are vast and include both academia and companies and organizations around the world (Mora, 2019). Data science is a multidisciplinary field that combines knowledge from mathematics, statistics, computer science, and domain expertise to extract meaningful information from data. This process involves collecting, cleaning, analyzing, and interpreting data to make informed decisions. Some of the applications are as follows:

- **Data Analysis:** Use statistical techniques and machine learning algorithms to identify patterns and trends in the data.
- **Machine Learning:** Develop predictive models and classifiers based on historical data to make predictions about new cases.
- **Natural Language Processing:** Extracting meaningful information from text, such as

“Throughout this chapter, we will focus on the potential and applications of AI to develop data analytics”

sentiment analysis, machine translation, and article summarization.

Throughout this chapter, we will focus on the potential and applications of AI to develop data analytics. From the most basic concepts to the most advanced, we will introduce AI applications in spreadsheets, continue with an exploration of AI capabilities in advanced statistics, and finally present some opportunities in machine learning.

9.1. Within spreadsheets

Spreadsheets (Microsoft Excel, Numbers, Google Sheets) are probably one of the most popular and versatile tools for performing more or less sophisticated operations on data organized in rows and columns. Their applications and uses are very diverse and have different levels of complexity. Some spreadsheets are simple records of ordered information, but these tools allow for more complex processes, such as evaluations, operational processes, decision making, financial reporting, or data analysis (O’Beirne, 2023). The popularity of these tools has generated an excess of confidence, which O’Beirne calls the “sorcerer’s apprentice syndrome,” who has enough knowledge to make an enchantment, but not what is needed when it gets out of control.

AI makes it easier to use spreadsheets, as generation tools such as Copilot, ChatGPT, Gemini, or Claude can suggest precise instructions to achieve a

specific purpose, especially when the data is already available. Excel’s Data Analyzer can also be used as a built-in assistant the software. Finally, AI assistants (ChatGPT, Copilot) can be used within Excel or Google Sheets. We will show examples of these three possibilities.

Using AI to Guide and Execute Analysis

Artificial intelligence (AI) tools can be invaluable for generating ideas and supporting the data analysis process. Indeed, as evidenced in the preceding sections, tools such as ChatGPT and Copilot are invaluable in addressing queries or resolving particular scenarios or issues. To illustrate, consider the following dataset, represented in an Excel spreadsheet, comprising the physical characteristics of 20 individuals (age, height, sex, weight, and level of physical activity).

CASE	AGE	HEIGHT (cm)	SEX	WEIGHT(kg)	Physical activity level
1	25	170	M	70	Moderate
2	32	165	F	60	Low
3	28	180	M	80	High
4	45	160	F	55	Moderate
5	22	175	M	75	Low
6	30	168	F	65	High
7	35	172	M	77	Moderate
8	40	158	F	58	Low
9	27	181	M	82	High
10	26	164	F	62	Moderate
11	38	162	M	61	Low
12	32	174	F	73	High
13	24	160	M	54	Moderate
14	31	179	F	81	Low
15	29	164	M	69	High
16	17	169	F	52	Moderate
17	23	165	M	66	Low
18	27	180	F	79	High
19	20	177	M	75	Moderat
20	28	163	F	57	Low

“Artificial intelligence tools can be extremely useful for generating ideas and supporting the data analysis process”

For example, AI can help us calculate the average age with the following prompt:

“I want to calculate the average age. Tell me the formula to use”

Its response would be the following:



To calculate the average age in Excel, you need to use the 'AVERAGE' function. If the ages are in column B from cell B2 to cell B21, the formula you need to write is:

```
excel Copia  
=AVERAGE ( B2 : B21 )
```

This will calculate the average of the ages of cells B2 to B21. Here are the detailed steps:

1. Select the cell where you want the average result (for example, cell B22).
2. Type the formula '=AVERAGE (B2:B21)'.
3. Press enter.

Excel will calculate and display the average of the ages in the selected cell.

Applying its instructions produces the following result:

CASE	AGE	HEIGHT (cm)	SEX	WEIGHT(kg)	Physical activity level
1	25	170	M	70	Moderate
2	32	165	F	60	Low
3	28	180	M	80	High
4	45	160	F	55	Moderate
5	22	175	M	75	Low
6	30	168	F	65	High
7	35	172	M	77	Moderate
8	40	158	F	58	Low
9	27	181	M	82	High
10	26	164	F	62	Moderate
11	33	177	M	78	Low
12	38	162	F	61	High
13	24	174	M	73	Moderate
14	29	166	F	59	Low
15	31	179	M	81	High
16	42	155	F	56	Moderate
17	23	169	M	74	Low
18	37	167	F	63	High
19	34	176	M	79	Moderat
20	28	163	F	57	Low
	31,45				


We could ask for a second instruction, such as the number of males and females in the sample. The prompt could be:

“I want to know the distribution of gender variables. Tell me the formula. Excel is in English”

The result of this query is as follows:

Count the number of female (“F”)


excel

 Copia el codi

```
=COUNTIF(D2:D21;"D")
```

Count the number of male (“M”)

excel

 Copia el codi

```
=COUNTIF(D2:D21;"H")
```

Thus, the sex data are in column D of cells D2 to D21:

1. **Number of females:** `=COUNTIF(D2:D21;"F")`
2. **Number of males:** `=COUNTIF(D2:D21;"M")`

The system has interpreted that M as male and F as female. In any case, the prompt can be refined

if the answer was not satisfactory. The result of the operation would be as follows:

CASE	AGE	HEIGHT (cm)	SEX	WEIGHT(kg)	Physical activity level
1	25	170	M	70	Moderate
2	32	165	F	60	Low
3	28	180	M	80	High
4	45	160	F	55	Moderate
5	22	175	M	75	Low
6	30	168	F	65	High
7	35	172	M	77	Moderate
8	40	158	F	58	Low
9	27	181	M	82	High
10	26	164	F	62	Moderate
11	33	177	M	78	Low
12	38	162	F	61	High
13	24	174	M	73	Moderate
14	29	166	F	59	Low
15	31	179	M	81	High
16	42	155	F	56	Moderate
17	23	169	M	74	Low
18	37	167	F	63	High
19	34	176	M	79	Moderat
20	28	163	F	57	Low
	Females		10		
	Males		10		

Once we understand the basic functionality, we can add more complexity to our commands, such as calculating a new variable based on the existing information. For example, we can ask how to calculate

the Body Mass Index (BMI) with the following prompt: "I want to calculate the Body Mass Index based on this data. Tell me how to do that".



To calculate the Body Mass Index (BMI) from the weight and height data, you can use the following formula:

$$IMC = \frac{pes}{(altura)^2}$$

Where weight is measured in kilograms (kg) and height in meters (m.) Here are the steps to calculate BMI in Excel:

1. **Create a new column for BMI:** Add a new column to your spreadsheet where you will calculate the BMI. Suppose you want to put the BMI in column F, start in cell F2.
2. **Apply the BMI formula:** to the first cell of the new column (for example, F2), type the following formula and press Enter:

excel

Copia el codi

```
=E2 / (C2/100)^2
```

This formula assumes that the weight is in column E and the height is in column C (and that the height is in centimeters, which is why it is divided by 100 when converted to meters).

3. **Drag the formula:** Drag the cell with the formula down to apply it to all cells in column F with data.

This will calculate the BMI for each person from your weight and height data. If you need any clarification or help, don't hesitate to ask.

If we apply its instructions, we will get this result in the spreadsheet:

CASE	AGE	HEIGHT (cm)	SEX	WEIGHT(kg)	Physical activity level	IMC
1	25	170	M	70	Moderate	24,22
2	32	165	F	60	Low	22,04
3	28	180	M	80	High	24,69
4	45	160	F	55	Moderate	21,48
5	22	175	M	75	Low	24,49
6	30	168	F	65	High	23,03
7	35	172	M	77	Moderate	26,03
8	40	158	F	58	Low	23,23
9	27	181	M	82	High	25,03
10	26	164	F	62	Moderate	23,05
11	33	177	M	78	Low	24,90
12	38	162	F	61	High	23,24
13	24	174	M	73	Moderate	24,11
14	29	166	F	59	Low	21,41
15	31	179	M	81	High	25,28
16	42	155	F	56	Moderate	23,31
17	23	169	M	74	Low	25,91
18	37	167	F	63	High	22,59
19	34	176	M	79	Moderat	25,50
20	28	163	F	57	Low	21,45

Generators like ChatGPT are powerful enough to perform automatic data analysis, such as descriptive statistical analysis, analysis of the relationship between variables, a graphical representation of information, or more advanced analysis. In the previous example, we could use the following prompt: "Perform a descriptive statistical analysis on this spreadsheet. Remember to attach the file containing the results

of this analysis." This prompt demonstrates a very versatile use of ChatGPT, as it can provide files in various formats that can be downloaded; these files are temporary and cannot be downloaded after a certain period of time. The prompt therefore generates a downloadable results table and a basic interpretation of the results.

	AGE	HEIGHT	WEIGHT	IMC
count	20	20	20	20
mean	31,45	169,05	68,25	23,75
std	6,50	7,56	9,52	1,48
min	22	155	55	21,41
25 %	26,75	163,75	59,75	22,92
50 %	30,5	168,5	67,5	23,71
75 %	35,5	175,25	77,25	24,93

1. General Statistics:

- Age: The average age is 31.45 years, with a range of 22 to 45 years.
- Height: The average height is 169.05 cm, with a range of 155 to 181 cm.
- Weight: The average is 68.25 kg, with a range of 55 to 82 kg.
- BMI: The median is 23.75, with a range of 21.41 to 26.03.

2. Gender Differences:

- Sex F (Female):
 - Age: The average age is 34.37 years.
 - Height: The average height is 162.3 cm.
 - Weight: The average is 61.5kg.
 - BMI: The average is 22.48.
- Sex M (Male):
 - Age: The average age is 28.2 years.
 - Height: The average height is 175.8 cm.
 - Weight: The average is 75 kg.
 - BMI: The average is 25.02.

These data show significant differences between males and females in terms of height, weight and BMI. Males tend to be taller and heavier, with a higher MFI compared to females.

We can see that the analysis of the results provides the mean and the range of all the values. In addition, it differentiates between the averages for the gender attribute and highlights the differences in values between men and women.

ChatGPT not only works with descriptive statistics but also allows for graphical representations. For example, we can use the following prompt: "Create a graph of

the data I have provided". Without any further input, we will see that it can show the relationship between variables. For the previous case, we could suggest the following prompt: "Calculate the correlations between variables, if possible." Again, it generates a downloadable table, with the correlation matrix and an interpretation of the results.

Age

- Moderately correlated with height (-0.558) and weight (-0.485)
- Has a weaker negative correlation with BMI (-0.318)

Height (cm)

- Very strongly positively correlated with weight (0.946)
- Moderately positively correlated with BMI (0.684)

Weight

- Strongly positively correlated with height (0.946)
- Strongly positively correlated with BMI (0.882)

BMI

- Strongly positively correlated with weight (0.882)
- Has a moderate positive correlation with height (0.684)

These correlations suggest that height and weight are strongly related, as are weight and BMI. Conversely, age is negatively correlated with height, weight, and BMI, although these correlations are weaker.

If you need more details or additional help, don't hesitate to ask!

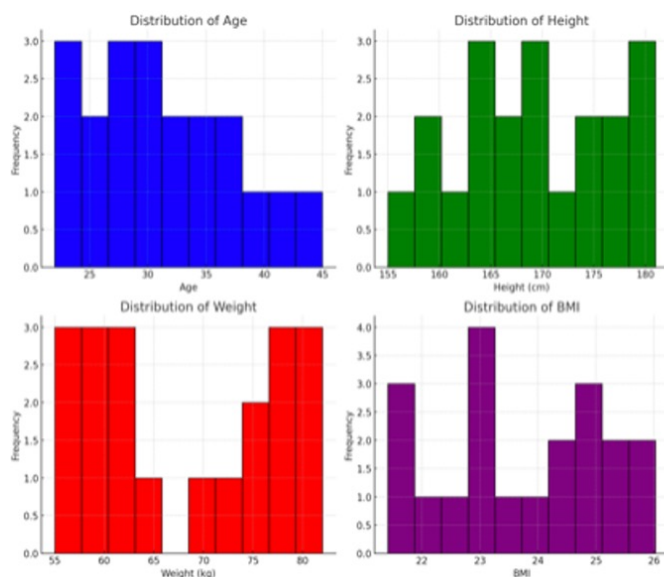
ChatGPT also allows for graphical representations. For example, we can suggest the following prompt: "Create a graph of the data I have provided." We will see that it can include the relationship between variables without further instructions. For example, in the previous case, we could suggest the following prompt: "Calculate the correlations between variables, if possible". Again, it will generate a downloadable

table with the correlation matrix and an interpretation of the results.

Without further instructions, the program will create a histogram for each variable studied and then provide scatter plots that allow us to visualize the relationship between variables.



Distribution Of BMI

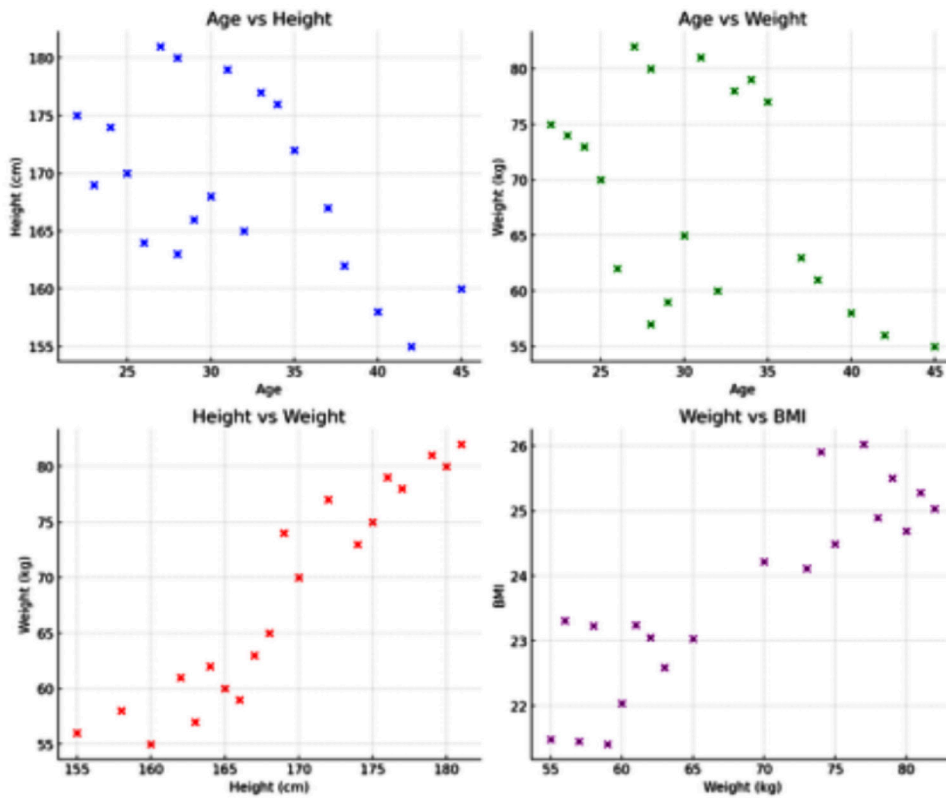


Here are the histograms illustrating the distribution of the main variables in the database:

1. **Age Distribution:** Most people are between 25 and 35 years old, with a dispersion of up to 45 years old.
2. **Height Distribution:** The height varies mainly between 160 and 180 cm.
3. **Weight Distribution:** weight is mainly concentrated between 55 and 80 kg.
4. **BMI Distribution:** The BMI varies between 21 and 26, with a concentration mostly around 23-24.

Now, I will create other graphs to show the relationships between variables, especially those that have significant correlations.

Weight Vs BMI



Using Assistants in Excel

Microsoft has introduced a tool in Excel called “Analyze Data” that allows users to get information about their data through natural language, allowing questions, reports, or suggestions without complicated formulas. In addition, this tool provides visual summaries, trends, and patterns (Microsoft, 2024). We entered the data

from the previous case into Excel and clicked on the “Analyze Data” option located on the left; this activates a menu that allows interaction with the spreadsheet and access to automatically generated graphical information, as shown in the figure.

The screenshot shows an Excel spreadsheet with the following data:

Case	Age	Height (cm)	Sex	Weight (kg)	Level of physical activity	BMI
1	25	170	M	70	Moderate	24,22
2	32	165	F	60	Low	22,04
3	28	180	M	80	High	24,69
4	45	160	F	55	Moderate	21,48
5	22	175	M	75	Low	24,49
6	30	168	F	65	High	23,03
7	35	172	M	77	Moderate	26,03
8	40	158	F	58	Low	23,23
9	27	181	M	82	High	25,03
10	26	164	F	62	Moderate	23,05
11	33	177	M	78	Low	24,90
12	38	162	F	61	High	23,24
13	24	174	M	73	Moderate	24,11
14	29	166	F	59	Low	21,41
15	31	179	M	81	High	25,28
16	42	155	F	56	Moderate	23,31
17	23	169	M	74	Low	25,91
18	37	167	F	63	High	22,59
19	34	176	M	79	Moderate	25,50
20	28	163	F	57	Low	21,45

Summary statistics from the sidebar:

- AGE AVERAGE:** 31,45
- SEX FREQUENCY:** Male (10), Female (10)
- Pivot Table:**

Sex	Moderate	High	Low	Grand T...
F	37,6667	35	32	34,7
M	29,5	28,7	26	28,2
Grand Total	33	31,8	30	31,45
- Scatter Plot:** 'Height (cm)' vs 'Age' showing a positive correlation.

In the automatically generated results, we can see a table that differentiates the results by gender and a scatter plot that shows the relationship between age and height. Like ChatGPT or Copilot, this tool also allows natural language queries. However, its operation is conceptually quite different, and we need to keep a few principles in mind. For example, prompts for this tool need to be very concise, and the names of the variables should be enclosed in single quotes (' '). It is also important to note that this tool cannot engage in a "conversation" like ChatGPT or similar tools, so each question must

be independent. It does, however, provide a list of recent inquiries or questions we've asked.

The images show the prompts required to calculate the average weight of individuals ("Average of Weight (kg)") or the average BMI by gender ("Average of Total BMI by Gender"). The system also provides suggestions and automatically generates prompts.

Analyse Data

← average of "height (cm)"

average of "height (cm)"

Here's our answer
Showing average 'Height (cm)'.

'Height (cm)'	
Average of Height (cm)	
	169,05

[+ Insert PivotTable](#) Is this helpful?

Analyse Data

← average of total "BMI" by "sex"

average of total "BMI" by "sex"

Here's our answer
Showing total and average 'BMI' by 'Sex'.

'BMI' by 'Sex'		
Sex	Average of BMI	Sum of BMI
F	22,48	224,84
M	25,02	250,16
Grand Total	23,75	475,01

[+ Insert PivotTable](#) Is this helpful?

On the other hand, it is also possible to install a ChatGPT or Copilot assistant in both Excel and Google Sheets. Its functionality is quite similar to what we have seen with the "Analyze Data" feature integrated within Excel, but much more versatile, as it allows the use of natural language in a way that more closely resembles how we would use it directly in the ChatGPT or Copilot environment.

9.2. Advanced Statistics

One of the functionalities of ChatGPT is the creation of assistants or GPTs with specific functionalities. Subsequently, this tool's potential applications will be discussed in greater detail. Among these assistants, some are specialized in advanced statistical analysis. In order to illustrate the capabilities of these assistants, we will utilize two of the most powerful ones such as Data Analysis & Report AI and Data Analysis - SPSS. Since the assistants are frequently modified and updated with more advanced versions, it is possible that by the time you read this document, new GPTs with additional capabilities may have been introduced. ChatGPT has a search function that allows users to easily find specialized assistants in a given field.



Data Analysis & Report AI

De: Kenneth Bastian &

Your expert in limitless, detailed scientific data analysis and reporting.

Can you analyze this dataset for patterns?

What trends can you identify in this data?

Please generate a report based on this data.

Can you create visualizations for this dataset?



Data Analysis - SPSS

De: Meddidin ebuseerfo &

Expert in statistical analysis and SPSS, guiding researchers.

How do I test the reliability of my data in SPSS?

What statistical test should I use for my hypothesis?

Can you explain how to define variables in SPSS?

I need help interpreting SPSS output for my data.

Data Analysis & Report AI provides information in the Python language, unlike the models we have seen so far, which offer the possibility of downloading result in tables or graphs. This assistant is particularly useful for complex analyses, which it can illustrate with concrete examples. We can use the following prompt based on the data we use throughout this

chapter: "What types of nonparametric analyses can we perform on this data?" The result will show both a list of the most appropriate tests that could be performed and the downloadable Python code that will allow us to automatically run the selected option.



Non-parametric analysis is useful when the data do not follow a normal distribution or when we have small samples. In the context of these data, we might consider various nonparametric tests to compare groups and explore relationships:

1. Mann-Whitney U test 2

This test is used to compare two independent samples. It could be useful to compare differences between men and women (e.g., height, weight, or BMI).

2. Kruskal-Wallis test

This test is an extension of the Mann-Whitney U test for more than two groups. It can be used to compare differences in weight, height or BMI between different levels of physical activity (Low, Moderate, High).

3. Wilcoxon test

This test is used to compare two related samples. It may be applicable if we have data before and after an intervention for the same people (although this does not seem to be the case with the current data).

4. Spearman's Correlation

This test is used to assess the strength and direction of the association between two continuous or ordinal variables. It could be used to explore the relationship between age and BMI, height and weight, etc.

Concrete Examples

1. Mann-Whitney U Test: Comparison of Weight between Men and Women

```
python Copy the code  
  
from scipy.stats import mannwhitneyu  
  
# Split the data by sex  
men = df[df['Sex'] == 'M']['Weight (kg)'] women = df[df['Sex'] == 'D']['Weight  
(kg)']  
  
# Take the test  
stat, p = mannwhitneyu (men, women)  
state, p
```

2. Kruskal-Wallis Test: Comparison of BMI according to Physical Activity Level

```
python Copy the code  
  
from scipy.stats import kruskal  
  
# Split the data by physical activity level  
low = df[df['Physical activity level'] == 'Low']['BMI'] moderate = df[df['Physical activity level']  
== 'Moderate']['BMI'] high = df[df['Physical activity level'] == 'High']['BMI']  
  
# Take the test  
stat, p = kruskal (low, moderate, high)  
state, p
```

3. Spearman Correlation: Relationship between Age and BMI

```
python Copy the code  
  
from scipy.stats import spearmanr  
  
# Perform the correlation test  
corr, p = spearmanr(df['Edad'], df['IMC'])  
corr, p
```

These are some of the nonparametric analyses that might be useful for exploring your data. If you would like me to perform any of these analyses or have any other preferences, please let me know!

The Data Analysis – SPSS assistant is a more specialized tool, as it provides guidance on performing data analysis using the SPSS software. If we pose the same problem as in the previous case and thus introduce the same prompt, we can proceed as follows: “What types of nonparametric analyses can we perform with this data?”. The outcome is distinct. On the one hand, it suggests other non-parametric tests that could prove useful for data

analysis, thereby establishing the two assistants as complementary. On the other hand, it provides the requisite information for carrying out the corresponding statistical analysis using SPSS. From an educational standpoint, this allows for greater complementarity with the teacher’s activity, either in conjunction to a flipped classroom model or as a means of providing support for the completion of complex exercises.



How to perform these analyses in SPSS:

Mann-Whitney U test:

1. Go to Analyze > Nonparametric Tests > Legacy Dialogs > 2 Independent Samples.
2. Move the dependent variable to the Test Variable List box and the independent variable (groups) to the Grouping Variable box.
3. Define the groups by clicking Define groups.
4. Select Mann-Whitney U and click OK.

Wilcoxon test:

1. Go to Analyze > Nonparametric Tests > Legacy Dialogs > 2 Related Samples.
2. Move the two variables to be compared to the Test Pairs box.
2. Select Wilcoxon and click on OK.

Kruskal-Wallis test:

1. Go to Analyze > Nonparametric Tests > Legacy Dialogs > K Independent Samples.
2. Move the dependent variable to the Test Variable List box and the independent variable (groups) to the Grouping Variable box.
3. Define the groups by clicking on Define Range.
4. Select Kruskal-Wallis and click OK.

Friedman's test:

1. Go to Analyze > Nonparametric Tests > Legacy Dialogs > K Related Samples.
2. Move the variables to be compared to the Test Variables box.
3. Select Friedman and click OK.

9.3. Machines that Learn: Machine Learning (ML)

Machine learning (ML) is a subfield of artificial intelligence that postulates that computers can learn from patterns and experience, just as humans do. This enables them to make decisions based on previously acquired experience (Zhou, 2021). Consequently, machine learning (ML) is a technique that improves system performance through automatic learning using computational methods (Zhou, 2021). The development of statistical models and algorithms enables computers to learn from the data they are given and make predictions or decisions with a certain degree of autonomy (Rodríguez, 2023). In essence, the aforementioned processes enable computers to learn from the data they are given and make predictions or decisions with a certain degree of autonomy.

In recent years, machine learning has made significant progress due to the increase in computing power, the massive availability of data, and advances in the field of artificial intelligence. As a result, ML has become one of the most important techniques in applied areas such as computer vision and natural language processing (Zhou, 2021), playing an increasingly important role in supporting scientific research and being crucial for technological development (Mjolsness and DeCoste, 2001).

Today, many specialized software programs include algorithms and data analysis techniques based on ML. In fact, most major data analysis programs have ML tools or modules that facilitate the exploration of complex data. For example, the free software JASP includes an ML module with regression, classification, clustering, and prediction techniques. We also found software such as MAXQDA that allows for the analysis of textual data through natural language processing.

In this chapter, we will primarily focus on demonstrating two practical applications of ML. First, cluster analysis as a model for finding patterns in data; and second, an analysis based on data mining as a practical application of natural language processing.

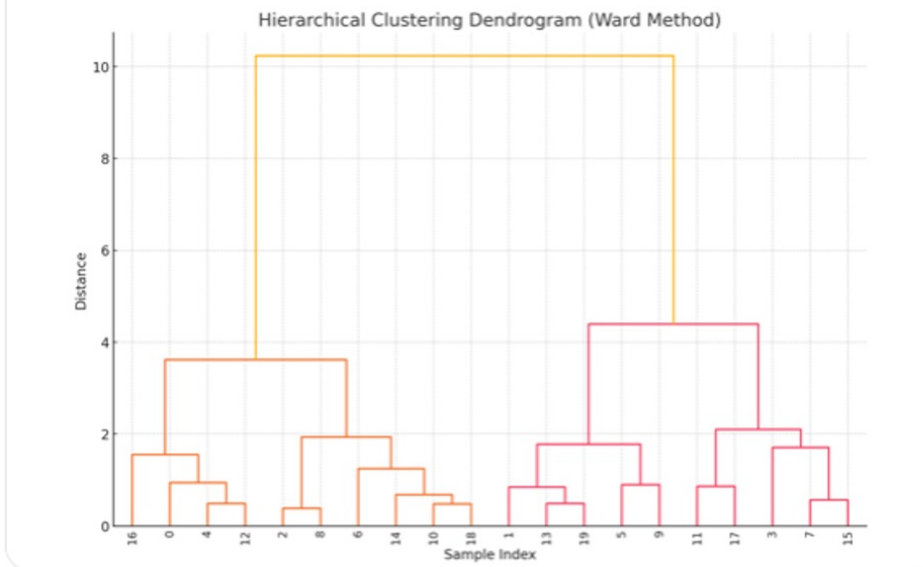
Pattern Search: Cluster Analysis

Cluster analysis is a statistical technique used to find hidden patterns or relationships in data by grouping elements that share similar characteristics. The data are grouped into a reduced number of mutually exclusive clusters or groups (Hair et al., 1998) (Cea, 2004). In short, several methodological decisions must be made when conducting a cluster analysis. The most common approach is to use the hierarchical method. There are different classification algorithms (inter-groups, intra-groups, ward method...), each with its own characteristics, which prioritize different connections within the data. They therefore emphasize different features in the data patterns. In addition, when dealing with variables measured on different scales, it is necessary to standardize the data (z-score) to avoid the potential influence of the squared Euclidean distance on the resulting clusters.

In the following example, we will use the same data we have been working with since the beginning of this chapter. Now we will ask ChatGPT to perform a cluster analysis using the Ward method clustering algorithm (since it is a small dataset) and to standardize the data with z-score. In this case, we will give explicit instructions on how to do it because we already have prior knowledge, but there is also the option of asking for suggestions on how to handle the data. So, we give the instruction with the following prompt: "Perform a cluster analysis on the table using the Ward method and z-score". The result is shown below, which includes the dendrogram, the data with the cluster classification, and the corresponding interpretation.



Hierarchical Clustering Dendrogram (Ward Method)



Here is the dendrogram resulting from the cluster analysis using Ward's method and z-score normalization:

- Each leaf of the dendrogram represents an individual in the data set.
- The height at which the individuals or groups are combined indicates the distance (or dissimilarity) between them. The greater the height, the more dissimilar the groups are.
- The dendrogram helps to identify the name of natural clusters within the data set. You can cut the dendrogram at different levels to obtain different cluster names.

If you need a more detailed interpretation of the clusters or want to cut the dendrogram at a specific level to get a specific number of clusters, let me know.

Yes, you must interpret the results by naming the clusters and indicating the best solution.



Clustered Data					↓	↗
	Case	Age	Height (cm)	Sex		
1	1	25	170	H		
2	2	32	165	D		
3	3	28	180	H		
4	4	45	160	D		
5	5	22	175	H		
6	6	30	168	D		

The cluster analysis resulted in two main groupings:

1. Cluster 1:

- Average age: 28.2 years
- Average height: 175.3 cm
- Average weight: 76.9 kg
- Average BMI: 25.02

2. Cluster 2:

- Average age: 34.7 years old
- Average height: 162.8 cm
- Average weight: 59.6 kg
- Average BMI: 22.48

Cluster interpretation:

• Cluster 1 (Tall young people):

- This group is made up of younger people, with a greater height and a higher weight. Their BMI is also higher compared to the other group.

• Cluster 2 (Low Height Adults):

- This group is made up of slightly older people, with a lower height and lower weight. Their BMI is lower compared to the other group.

Best solution:

- The best solution is to divide the population into two clusters reflecting significant differences in physical characteristics (height, weight) and age. These two clusters are easily interpretable and show clear differences between the two groups of people.

If you need more details or further analysis, please let me know.

This is an example of the potential of performing data analysis using ML techniques through ChatGPT. However, it is necessary to verify that the result is logical and correct according to the data. This can be a way to explore the data before using statistical software, such as JASP to perform the final analysis.

Natural Language Processing: Text Mining and Sentiment Analysis

The accumulation of large amounts of textual data, such as in social networks, websites, and blogs, has increased the need for efficient methods to process, manage, and analyze this data. Therefore, techniques and tools are needed to facilitate the extraction of relevant information from this data so that it can be easily interpreted. Text mining is the set of techniques and methods used to analyze such data and enable the extraction of meaningful knowledge and patterns (Gaikwad, Chaugule, and Patil, 2014). The field includes various processes such as data cleaning, transformation, and modeling, with the aim of obtaining relevant information.

To perform this analysis, various algorithms and statistical models are used to examine textual data in detail. These models are able to handle the complexity and variability inherent in textual data and help to identify useful patterns and trends that define significant features within the analyzed data (Lau, Lee, and Ho, 2005). For example, identifying recurring words or themes can provide valuable insights into users' preferences and interests.

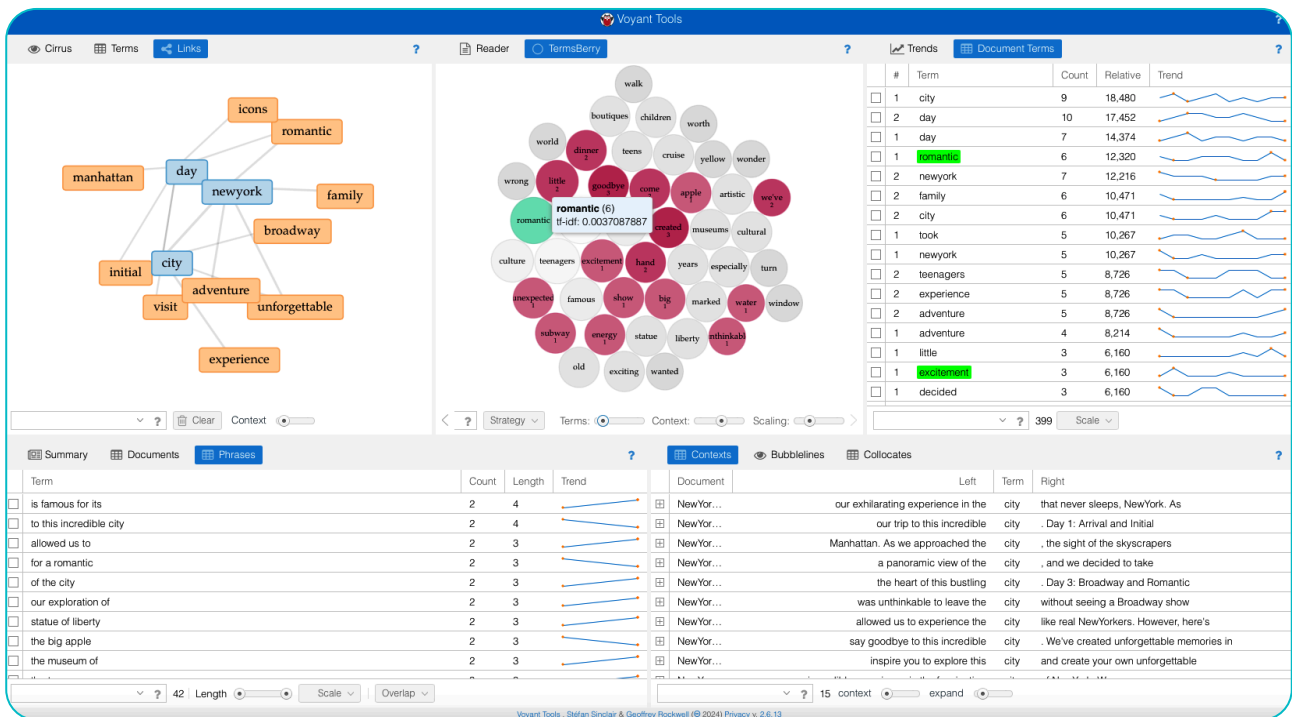
Among the more specific techniques employed in the field of text mining are sentiment analysis and opinion mining. These techniques permit the determination of an individual's attitude and feelings toward a particular product or subject matter, thereby furnishing crucial information for businesses and researchers. Sentiment analysis, for instance, can classify opinions as positive, negative, or neutral, while opinion mining can identify specific topics that are important to users (Khan, Qamar, and Bashir, 2016). This type of analysis is particularly beneficial in the context of digital marketing, market research, and online reputation management, where

an understanding of customer opinions can have a significant impact on the business strategy adopted.

The application example we will see next uses Voyant Tools (<https://voyant-tools.org>). It is a web-accessible application that allows you to perform textual data analysis based on the principles of text mining. This is not the only tool available; there are many others, both online and offline, that operate and work according to the principles of data mining.

Suppose we want to analyze the image of New York through travel blogs. To proceed with the analysis, we need to have all the blog texts in a file. In the example we will present, we have two texts in two separate files. From there, the texts need to be uploaded to the tool (Voyant Tools) to ensure that all the texts are properly formatted for analysis. This tool allows for the execution and retrieval of all results in a visual and easy-to-understand format. The results we obtained are presented below:





We can see how information is obtained about the most frequent words, the different themes, and the thematic relationships that emerge from the analyzed texts, as well as the frequency of phrases and words, etc. This tool has many applications in teaching: its speed and accessibility allow the class to focus on analyzing the results or comparing different cases. However, it is crucial that students clearly understand the statistical and operational processes behind content analysis in order to better

interpret the results and recognize the limitations of the models. The possibilities in the field of social sciences are almost limitless: studies of political discourse, analysis of conversations based on documentary transcripts, collection and analysis of tweets related to racism, semantic analysis of media based on their orientation, analysis of reggaeton lyrics, or study of tourist descriptions of destinations.

Activity. “How do you spend your free time?”

Objective

- Explore the possibilities of data analysis with ChatGPT.
- Work with open databases.

Considerations

- Aimed at Social Sciences students.
- Activity is best done in groups.
- AI tool: ChatGPT or Copilot
- The exercise will be carried out with an open database, such as those provided by the Statistical Institute of Catalonia (<https://www.idescat.cat/dades/micro/>). The most versatile ones are the Living Conditions Survey, the Time Use Survey, and the Linguistic Habits Survey, which provide access to microdata.

For this activity, we will primarily work with the Time Use Survey database, which provides insights into the time spent on different activities (personal care, paid work, study, unpaid work, household and family, leisure activities, and commuting) by a representative sample of Catalans. Given the large amount of information and the high number of variables, it is recommended to focus on a specific area, such as the inequality of time spent on domestic activities by each gender.

Once the study variable(s) have been selected, the following tasks are required:

- Descriptive statistics (frequency tables, mean, standard deviation, minimum, and maximum, etc.)
- Explore possible relationships between different variables in the data set.
- Create graphical representations of the data as deemed appropriate.
- Write a short report interpreting the data.

Upon completion, share with the class group:

- Share reflections and results obtained.
- Draw general conclusions about the data analysis performed.
- Discuss the main advantages and challenges of using AI tools for data analysis.

Activity. The Conversation about COVID-19

Objective

- Learn to use AI tools to effectively analyze textual data.
- Develop skills in interpreting and communicating the results of text analysis.
- Become familiar with textual data sets.

Considerations

- Intended for students in the Social Sciences and Humanities.
- Activity is best done in groups.
- AI tool: Voyant Tools
- Dataset: [kaggle.com](https://www.kaggle.com).

There are many servers online that provide freely accessible text databases that are particularly useful for educational purposes. One of the platforms that provides access to a large amount of information is [kaggle.com](https://www.kaggle.com), a community focused on textual data with a large number of datasets; Kaggle is part of Google Cloud. In this case, we propose the possibility of accessing the COVID-19 tweets dataset (<https://www.kaggle.com/datasets/gpreda/covid19-tweets>), which contains a selection of tweets using this hashtag. There are many databases that can be adapted to the course, such as one dedicated to opinions about Biden or Trump, sentiments expressed in comments about the Olympic Games, movie reviews, or public debates on controversial topics. Voyant Tools (<https://voyant-tools.org/>) will be used to analyze the downloaded dataset. A subset of the tweets will need to be selected based on a specific criterion (temporal, geographic, thematic).

The suggested procedure is as follows:

- Upload the texts to Voyant Tools, making sure the texts are properly formatted for analysis.
- Develop a descriptive statistical analysis.
- Extract frequent words: Identify the most frequent words using the “Summary,” “Terms,” and “Document Terms” functions.
- Word distribution: Analyze the distribution of words throughout the text using the “Trends” tool.
- Graph and visualize data:
- Word visualization: Create a word cloud using the “Cirrus” tool to identify the most common words in the text.
- Visualize thematic relationships: Create a word network using the “Links” and “TermsBerry” tools to visualize the most common connections between words and themes.
- Write a short report and interpret the results obtained.

The following activities can be carried out in the class group:

- Gather the class group together to share their reflections and the results obtained.
- Establish group conclusions about the textual analysis performed.
- Discuss the main advantages and challenges of using AI to analyze textual data.

References

- Briones, J.A. (n.d.) <https://www.plenainclusion.org/wp-content/uploads/2022/02/Plena-inclusion-Murcia.-Guia-de-Excel.pdf>
- Gaikwad, S. V., Chaugule, A. y Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, 85 (17), 42-45.
- Khan, F. H., Qamar, U. y Bashir, S. (2016). eSAP : A decision support framework for enhanced sentiment analysis and polarity classification. *Information Sciences*, 367, 862-873.
- Lau, K. N., Lee, K. H. y Ho, Y. (2005). Text mining for the hotel industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46 (3), 344-362.
- Mannycia, James et al. (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. New York: McKinsey & Company [Consulted: 20/05/2024]. <http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>
- Microsoft (2024). Análisis de datos en Excel. [Consulted: 02/06/2024] <https://support.microsoft.com/es-es/office/análisis-de-datos-en-excel-3223aab8-f543-4fda-85ed-76bb0295ffc4>
- Mora, M. (2019). *Els rols, àmbits i noms de la ciència de dades*. FUOC.
- O'Beirne, P. (2023). Proceedings of the EuSpRIG 2023 Conference "The Spreadsheet Crisis: Regaining Control". European Spreadsheet Risks Interest Group.

[1] This prompt may include an indication of the Excel's language, as different versions of the program come with translated formulas. In this example, Excel is in the English version.

10

Language models in qualitative analysis

Qualitative research methodologies play a fundamental role in the field of social sciences, offering a deep and nuanced understanding of human behavior and the motivations that drive it. Through data collection methods such as interviews and focus groups, researchers can gather rich, detailed information that sheds light on why people think and behave in certain ways. Epistemologically, qualitative research does not reveal a central 'truth' that can be generalized but rather produces transferable results (Maxwell, 2013), which appear valid to the participants in the study, based on their expressions. Through a careful process of coding and interpreting non-numerical data, nuanced narratives are transformed into understandable themes, and meaningful patterns and insights are then constructed (Burnard, 1991; Jalali & Akhavan, 2024).

To achieve this, robust data collection methods, appropriate informant participation and engagement, meticulous coding, and triangulation, among others, are used to enhance reliability. This process is inherently long and labor-intensive. Because of this, technological advances in software packages, including ATLAS.

ti, NVivo, and MAXQDA, as well as natural language processing approaches such as sentiment analysis and topic modeling, assist researchers in producing useful analytical material while managing and interpreting large volumes of textual data (Zhang et al., 2024).

As a result, in recent years, artificial intelligence (AI) transcription tools with sophisticated voice recognition algorithms have been developed, automatically translating audio and video recordings into written text, allowing researchers to focus their attention on analysis (Christou, 2023). Additionally, AI language tools such as Claude, ChatGPT, or Gemini have been incorporated into qualitative research. Using deep learning techniques and the ability to process and generate human language, they can perform tasks ranging from simple text generation to thematic categorization and contextual understanding (Xiao et al., 2023). Specifically, ChatGPT has been studied as a useful tool in qualitative research due to its ability to add a new level of depth to data analysis, as it understands natural language and allows users to easily interact with it. It is resistant to user input errors, such as

“During data analysis, it has been found that AI language tools like ChatGPT or Gemini tend to be more descriptive than interpretive”

spelling or grammatical mistakes, responds efficiently to requests and data interpretation processes, and provides explanations and details about the results, offering transparency (Zhang et al., 2024). As a tool, it goes beyond capturing limited information like phrase frequency or emotional tones (through sentiment or content analysis) and extends to identifying political, social, or cultural issues, enhancing the researcher's ability to refine codes, build conceptual connections, and create theory (Anis & French, 2023). In response, the previously mentioned software packages (ATLAS.ti, NVivo, and MAXQDA) have developed collaborations with OpenAI, improving the coding process in terms of time and analysis.

- Although the use of AI in qualitative methodologies offers an interactive and time-saving workflow, there are some challenges to consider (Anis & French, 2023; Christou, 2023; Morgan, 2023; Jalali & Akhavan, 2024; Zhang et al., 2024). During data analysis, it has been found that AI language tools like ChatGPT or Gemini tend to be more descriptive than interpretive.
- AI lacks common sense and a contextual understanding of social and cultural elements, especially when ambiguous issues are presented. These outcomes are often referred to as 'hallucinations,' which are essentially produced by the limitations of context (previous queries and response content) as the AI attempts to predict what might come next. Moreover, AI may be unable to comprehend human-generated text that includes metaphors or sarcasm.
- Another challenge in using AI systems is their tendency to oversimplify, given that they are trained to prioritize efficiency and simplicity over complexity of environments and phenomena. This could lead to the loss of valuable insights. As discussed in a previous chapter, AI language tools carry inherent biases because they are trained on human-generated data. Despite efforts to limit these biases, it is expected that the results will reflect dominant ideas and beliefs prevalent in society. This could affect the validity and reliability of the research, especially if the researcher does not carefully manage the data analysis process.
- AI should enhance researchers' capabilities rather than replace them. While it can assist with data interaction, it cannot be considered an independent or objective interpreter of data, much less a theory-building machine. Qualitative research embraces subjectivity, and it is here that researchers' knowledge, understanding, and perspectives are critical for reflecting on and interpreting data, particularly in ethnographic or phenomenological studies. It is the responsibility of researchers to question, refine, and include or exclude the codes, themes, and patterns produced by AI.
- Ethical and data ownership concerns are growing in the field of human-based data sharing, including personally identifiable information or private stories derived from personal interviews or focus groups, as there is no clear guidance on how these should be used or shared. This is often the case

with automated transcription programs, which also have the drawback of reducing researchers' opportunity to engage more deeply with qualitative data. To avoid ethical concerns, personal data must be removed before being entered into the programs, and consent forms must be obtained from participants regarding data protection.

Qualitative researchers continue to experiment with AI tools despite these challenges, in order to demonstrate their usefulness as producers of materials and as assistants in research. Future innovations may enhance AI's ability to understand more nuanced datasets and provide deeper qualitative analysis. In the field of textual data coding and thematic analysis, articles on this topic contribute to enriching collaborations between researchers' expertise and AI tool responses.

10.1. Assistance with the question guide

However, qualitative methodologies require data generation before analysis. Most research related to AI tools and qualitative research begins with researchers already having rich textual data from interviews; typically, the result of semi-structured interviews or focus groups. However, the quality of data collected during an interview is not solely determined by the interview protocol quality. Particularly in qualitative methodologies, data generation is a complex process that requires, in addition to prior study and preparation of the interview protocol, continuous attention and adaptation to the interviewees' inputs to ensure that all interesting data, whether explicitly or implicitly expressed, are adequately captured and described.

As researchers, students must decide whether to act as mere listeners during an interview or as co-creators of data. The latter is particularly essential in semi-structured or unstructured interviews. First, students must prepare a solid interview protocol. A common mistake is to jump immediately into the questions they wish to ask without breaking the ice, without learning enough about their interviewees,

and without connecting the theories previously read with the empirical case being studied. Preparing a comprehensive protocol requires significant time, which is where AI language tools like ChatGPT can be useful. To formulate appropriate questions based on demographic or psychographic characteristics, it is essential to clearly outline the research objective in the prompt. Since semi-structured interviews are the most common, it is beneficial to indicate the number of questions you intend to ask (around 30) and their format (open-ended). Based on what students have read in the literature review, it is worthwhile to revise the questions and add any missing concepts regarding the background, experiences, and perceptions of the interviewees.

For example, when we ask ChatGPT (version 4.0) to create an interview protocol under this objective: *'To explore and understand the experiences, challenges, and lifestyle adaptations of female digital nomads in urban and rural environments. This study aims to uncover how the physical and social dynamics of these settings influence their work-life balance, professional productivity, and personal well-being. By examining the nuances of their interactions with different cultural and infrastructural landscapes, the research intends to contribute to the broader knowledge of remote work and gender in the digital economy,'* the following elements were identified as subject to modification:

- Only the first question (*'Can you briefly describe your professional background?'*) asked about the interviewees' profiles, while the rest were directly related to the digital nomad lifestyle. More questions should be added to better understand the interviewee (personal life, previous jobs, travel experiences) and what life was like before making this decision.
- Some questions need to be simplified when they contain too many variables and may be difficult to recall. For example, *'Do you find certain types of work easier or harder in urban or rural environments?'* Here, ChatGPT created a question that asks for both negative and positive elements, as well as

“Qualitative researchers continue to experiment with AI tools despite these challenges, in order to demonstrate their usefulness as producers of materials and as assistants in research”

urban and rural settings. It should be split into at least two questions.

- Certain questions guide the type of response provided. For example, *‘Have you faced any situations of professional isolation? How do you manage them?’* is a question that directs the conversation toward the particular problem of isolation, potentially overlooking other situations more relevant to the interviewee. Rephrasing this question to be open-ended can allow for a broader range of responses (*‘What are the professional challenges this lifestyle has brought you? How have you managed them?’*).

If you want to connect this protocol with a specific literature review, you can incorporate the most

relevant articles into the prompt and ask the generator to add some questions to the protocol based on these articles. Afterward, you can select and assess the proposed questions. For example, in the case mentioned, ChatGPT was asked to consider an [article](#) on the topic and generate questions related to the gender perspective in the digital nomad lifestyle. This resulted in the creation of ten questions related to the article’s content, two of which were removed because they were too similar to the original protocol. In this way, an interview guide of about 40 questions was created, which can be very useful for a student with little experience developing interview scripts.

Activity: the interview assistant

Ask ChatGPT to design a semi-structured interview protocol, with approximately 30 open-ended questions, based on the following objective:

Explore and understand the motivations behind international students' decisions to pursue undergraduate programs at universities across Europe, examine their experiences during the mobility period, and assess the impact of these experiences on their personal development and personality traits. This research will investigate the main factors that influence international students' decisions, including academic, social, cultural, and economic motivations. Additionally, it will examine the lived experiences of these students, evaluating their adaptation to different educational systems, social environments, and cultural contexts, as well as the challenges they face and the support systems they rely upon. Additionally, the study aims to assess the changes in personality traits and personal development resulting from their international mobility, understanding the impact of intercultural interactions, academic challenges, and increased independence on their self-perception and growth. Finally, the research will compare students' pre-mobility expectations with their post-mobility reflections to analyze how their real experiences align with or differ from their initial motivations and expectations.

- a) Ask students to review the generated questions to:
 - Obtain the necessary profile information for international students
 - Ensure that each question addresses only one variable to avoid confusing the interviewees
 - Ensure that the questions do not suggest a particular response
 - Convert closed-ended questions into open-ended ones
 - Add and highlight questions that were not included in the proposed protocol but that you consider relevant to ask.
- b) Ask the students to read an article related to International Student Mobility (for example: Tran, L. T. (2016). Mobility as 'becoming': A Bourdieuan analysis of the factors shaping international student mobility. *British Journal of Sociology of Education*, 37(8), 1268-1289). After reading it, they should suggest new questions based on the information they have learned and justify their importance.
- c) Students should add this article to ChatGPT, in the same chat where the protocol was generated, and ask it to suggest new questions based on the information contained in the article. Then, they should compare these questions with the ones they had previously suggested and have a discussion about the contribution of AI to generating the protocol.

Objective

Enable students to effectively leverage AI to generate initial interview questions and structure the protocol while critically evaluating and refining these questions to ensure they align with the research objectives, ethical standards, and the complexities of the topic being studied. This task aims to enhance students' ability to develop insightful and relevant interview questions related to prior research and to prepare an interview protocol that provides them with confidence that all potential discussion topics are covered (although new questions will be added after the first interviews).

In qualitative research, it has been shown that in order for interviewees to disclose personal information and describe in detail their perspectives, perceptions, and experiences, trust must be built between

them and the researcher. In semi-structured and unstructured interviews, success lies in the fact that the questions following each response are based on the content of that response. Interviewers must be trained to listen attentively to all elements of each answer and automatically formulate new open-ended clarification questions based on unexpected reasoning or facts provided by the interviewees. The result of this process will be the generation of complex and in-depth data, which must be carefully coded to enrich data analysis.

There are two basic techniques used in interviews that allow for a deeper exploration of interviewees' responses: the laddering technique and bracketing. Laddering is a qualitative research technique used to uncover individuals' motivations, values, and beliefs that influence their decisions and behaviors (Miles & Rowe, 2004). During this process, the interviewer focuses on specific attributes of a product or service and asks progressively deeper questions to reveal the associated consequences and, ultimately, the individual's core values. Through the Attributes-Consequences-Values (ACV) model, the interviewer moves from tangible characteristics to the benefits these characteristics provide to the core values they satisfy. This approach offers deep insight into consumer behavior, informing strategic decisions in marketing, product design, and user experience by connecting superficial preferences with deeper psychological drivers. Essentially, this is achieved by asking why the interviewee considers certain attributes and consequences important.

In contrast, bracketing refers to the methods researchers use to keep the interview personal, even when they already know a lot about the topic being discussed (Sorsa et al., 2015). Thus, students suspend their own assumptions about the world, allowing participants to describe their point of view, develop subjectivity without bias, and build trust between participants. For example, an interviewer may have an understanding of what a participant in a study means when they declare themselves to be a compulsive online shopper, but more questions must be asked to allow the participants to describe the situations in which these behaviors occurred, providing more detailed information.

This is where AI comes into play to train students. For trust to be created and for the interviewee to engage in a conversation rather than an interview, the student must automatically formulate appropriate follow-up questions at the right time, regardless of any prepared protocol. The "conversational" ability of ChatGPT, for example, allows users to interact with it through natural language (Zhang et al., 2024). In this context, ChatGPT can simulate unstructured interviews by acting as the interviewee after the research objective and interviewee profile have been described in the prompt for ChatGPT to portray.

Activity: The Language Model as a Training Partner

In this exercise, students will interact with Gemini, who will play the role of a specific person. Through this exercise, they will practice generating spontaneous and relevant questions based on the responses they receive, dynamically adapting their inquiry to gain deep insights into the subject's experiences and perspectives. Gemini will assume the role of a 32-year-old single mother living in a small apartment near Retiro Park in Madrid. She works as a digital marketing executive at L'Oreal and has a 7-year-old daughter. The interview will explore how she organizes vacations with her daughter, focusing on the challenges she faces, her decision-making process, and the options she considers, providing a nuanced understanding of her experiences balancing personal life and work responsibilities.

Instructions

Start the Interview: Begin by introducing yourself and explaining the purpose of the interview. Open the interview with the following questions:

"Thank you for taking the time to speak with me today. To start, could you tell me about the last time you planned a vacation with your daughter? What were some of the key factors you considered when choosing your destination?"

Conduct the Interview using the Laddering and Bracketing Techniques:

- Allow the conversation to flow naturally.
- Focus on open-ended questions to encourage detailed responses.
- Pay attention to her responses and formulate your next questions based on the information she provides.
- Explore not only the "what" but also the "how" and "why" behind her vacation planning processes and decisions.

General Topics to Explore:

- Her priorities when choosing a vacation destination.
- How she balances work responsibilities with personal time.
- Financial and logistical considerations in vacation planning.
- Support systems or resources she may use.
- Challenges she faces as a single mother when planning vacations.

Reflection

After the interview, write a reflection on the process. Discuss how you adapted your questions based on the interviewee's responses and how you could improve these questions after reflecting on the interview.

Objective

The goal is to improve students' ability to conduct unstructured interviews, a key method in qualitative research. By interacting with a simulated interviewee created by Gemini, students will practice generating spontaneous and relevant questions based on real-time responses, dynamically adapting their inquiry and ensuring all information provided is well-explored.

10.2. Coding with a Language Model

Thematic analysis is a crucial tool in qualitative research due to its flexibility, depth, and systematic approach. It allows researchers to uncover and interpret the complex layers of qualitative data, providing valuable insights into human behavior, experiences, and social phenomena. It also helps identify, analyze, and report on patterns of meaning (themes) within the data. This process follows a six-phase approach: familiarization with the data set, data coding, initial theme generation, theme development and review, theme refinement and definition, and final reporting (Zhang et al., 2024). This methodology is time-consuming and requires significant manual effort, especially when dealing with large and complex data sets, which poses a challenge for novice researchers, as subjectivity plays a role in theme identification. For students, this challenge is even greater since they may be unfamiliar with the coding process.

Researchers have sought to demonstrate the potential usefulness of AI programs in qualitative data analysis by comparing their effectiveness to manual coding (Morgan, 2023; Jalali & Akhavan, 2024). The comparison between manual coding and AI-assisted coding, focusing on ChatGPT (version 4.0), shows that both can identify significant themes and feedback loops, though they may not always align. In other words, ChatGPT can provide part of the analysis that the researcher can consider when interpreting the data, enhancing the produced analysis. Notably, ChatGPT offers a more direct and objective approach, which can reduce the risk of introducing researchers' own biases. However, it has also been suggested that AI models perform better in deductive analysis rather than inductive analysis, making human involvement even more critical.

For students to understand that AI language tools are more useful as assistants in generating ideas and identifying data within thematic analysis—rather than as standalone data analysts—they

must observe the differences compared to manual coding. This means conducting a replication analysis, where interview transcripts that the students have produced are first manually coded and then coded using AI programs. The order is significant because AI language tools produce well-structured and seemingly complete text, which may appear sufficient to students unfamiliar with the complexity of theme identification. To understand the themes produced and evaluate their appropriateness based on personal subjective interpretations, familiarization with the data—i.e., reading the entire data set carefully—is a necessary first step.

Once familiarization with the data is achieved, the use of ChatGPT (version 4.0) can begin in identifying variables and links between them. At this point, it is crucial to construct appropriate prompts to effectively interpret the nuanced language. The quality of the interaction's results is closely related to the quality of the prompts. In this chapter, we will describe the basic strategies to follow when creating prompts for qualitative analysis (while a more detailed description of prompt generation is provided in an earlier chapter).

Based on Zhang et al. (2024), the provided indications should:

- Offer a detailed description of the task, that is, the nature of the data, the objective of the research, the expected outcomes, and any other considerations to remember (such as the format of the entries). For example, *“Here is a transcription of a focus group about live music consumption. The participants are 30 to 50 years old, have families and children, live in urban areas, and represent middle-class social backgrounds. Each paragraph represents a participant and starts with their name. Each paragraph starting with the same name is from the same person. The objective is to identify how music is consumed when one works full time and has children. Please read it first.”*
- Preferably use role-playing, asking ChatGPT to be an expert in qualitative analysis, although it is important to keep in mind that this strategy on its own is not enough.

- Define the methodology, that is, the objective of the task, which in the case of thematic analysis would be asking the program to find patterns and themes.
 - Specify the analytical process, that is, add instructions on how the AI should analyze the data. Indicate whether it should consider a specific theory (which must be described or attach another text that the program can read and follow) or generate the most recurring themes across all the data (a limit can be set, for example, up to ten), and then, based on the responses, choose those of interest and request a more detailed output. The more precise the nature of the recurring themes (for example, *“please provide the main themes that the mothers in this focus group mention as challenges when needing to attend a live music event compared to the main themes from the fathers”*), the more in-depth the analysis generated. Familiarization with the data allows students to have a general idea of specific issues that arose during the focus group, and they could start on this basis. It is important to note that the words “theme” or “subtheme” are better reserved for the later, more interpretive phase of the analysis, as themes are constructed at the end of this process.
 - Define the output format, for example, if the main recurring themes were requested, then ask for them to be clustered into broader themes and then presented separately with representative excerpts and a brief summary. You can request a bullet-point list, textual format, or a table, and a shorter or longer description of the data. As for the content, benefits, drawbacks, and reasoning (reasons, suggestions, perceptions) can be requested separately. It is recommended to be as specific as possible in the method of analysis, for example, *“I want to identify the causal links between variables. You can consider all the previously identified variables and if the relationship between two variables comes from different parts of the focus group. Keep in mind that you can always go back and read the focus group data if you feel something is missing.”*
 - Include prioritization when requesting, for example, a code of ten sub-codes, starting with the most frequent or the most relevant to key objectives. For example, *“Describe the top seven disadvantages participants cite for attending festivals, starting with those caused by external factors and ending with those caused by individual reasons.”*
 - Explain the objectives of the analysis by asking ChatGPT to explain its decision-making process when choosing the results produced. The more analytical the requested output format, the easier it will be to offer transparency. Additionally, as with manual coding, analyzing each line or paragraph of data is more effective for in-depth analyses compared to analyzing the entire data set at once. The process takes more time, but the results are also more reliable.
 - Acknowledge its expertise to confirm to the program that the way it proceeds with the requested analysis meets the researcher’s expectations.
- Given the complexity of data analysis within qualitative methodologies, it is not surprising that pre-existing coding tools, such as ATLAS.ti, N-Vivo, and MAXQDA, have incorporated AI into their software. In particular, the oldest of these, ATLAS.ti, was launched in 1993, helping researchers carry out grounded theory and content analysis. Thirty years later, in 2023, it launched a beta version with AI, offering a fast alternative resource for text analysis, especially coding. Researchers can use this AI-based coding tool to process data by, on the one hand, visualizing the results and, on the other, predicting which codes or labels might be useful to apply to their dataset.
- Upon creating a new project and adding the interview or focus group transcripts to the ATLAS.ti software, automatic coding begins with a click on “Analysis.” It is important to select all the documents (transcripts) you wish to code, in order to create a common codebook. By clicking on “Analysis,” one of the available options

is “AI coding,” which refers to an emergent coding of your data, creating categories, codes, and subcodes. Here, there is no guideline on what should be coded, so a large number of codes will be created. Researchers must thoroughly review this list, as many codes may be repeated with slightly different wording, such as “age,” “aging,” or they may disagree with the codes assigned to specific quotations because the content is not related to that code. In this case, these codes must be manually removed, and different codes should be assigned to these quotations. Once all the codes have been reviewed, you must click the “Apply” button to add these codes to the program and display a summary.

The AI works only on paragraphs (not sentences within the paragraph) and omits very short paragraphs. To achieve better quality and less chaotic coding, it is better to remove from the text (transcripts) any context unrelated to what you are interested in coding (but do not remove significant parts necessary for understanding the context). The software provides a comprehensive user manual for a step-by-step coding process. In addition to emergent coding, ATLAS.ti also offers “intentional AI coding” so that users have more control over the codes produced. As a result of this process, more precise coding can be generated,

based on basic questions related to our research objective. Once the documents to be coded have been selected, click again on “Analyze,” and ATLAS.ti will ask about the objectives, allowing the research question or hypothesis and any supporting context to be incorporated. Suppose we conducted interviews with international students at our university and asked them, “What motivated you to choose your degree and our university?” ATLAS.ti will generate specific questions based on this objective, along with relevant code category names, which can be reviewed and edited, removing questions that are not of interest or even adding your own questions. When you click “Start Coding,” ATLAS.ti will analyze the documents and gather coding suggestions. Once the coding is deemed correct, click “Apply,” and ATLAS.ti will apply the coding and display a summary.

Thus, there are two ways to use language models in teaching qualitative tools: they can be support instruments in the complex process of coding accumulated information, or they can help in gathering information and drafting interview scripts. In both cases, these tools are a support that facilitates enhanced versatility, but they should not be presented as an alternative to non-automated intervention.

Activity: Hybrid Coding

In this practice, students will code a transcript from an interview, preferably one they conducted themselves so they are familiar with the data. Alternatively, they may use the transcript from the interview conducted in the previous essay. In the context of thematic analysis, students must apply three different coding methods: manual coding, AI-assisted coding (ChatGPT), and AI coding using ATLAS.ti.

a. Manual Coding

Students will read the interview transcript to familiarize themselves with the data, identify and highlight key themes, categories, and subcategories, and develop a coding framework based on these themes. They will then manually code the transcript by assigning codes to different parts of the text and mapping relationships between various themes and categories. Finally, students should provide examples from the transcript to illustrate each theme and category.

b. AI-assisted Coding (ChatGPT Version 4.0)

Based on the guidelines mentioned earlier on how to create prompts, students must develop a set of prompts that replicate the questions and guidelines used during manual coding, ensuring they are clear and specific. After entering the transcript and prompts into ChatGPT, students will review and refine the output provided by ChatGPT to ensure it aligns with the thematic analysis approach.

c. AI Coding with ATLAS.ti

Using the ATLAS.ti user manual, students will upload the interview transcript in its original format to the software. To understand the differences between AI emergent coding and AI intentional coding, they should work on separate projects and request emergent coding in one and intentional coding in the other. In the first case, students will review the generated codes and categories, remove any misleading ones, and then present the final structure of the codebook. It is important to mention the new categories/codes that emerged compared to the two previous methods. In the second case of intentional coding, students should add the initial research questions they used for ChatGPT into the "intention" box in ATLAS.ti. Students will process the suggested questions and related generated codes, and then "start coding." It is important to compare the resulting codes with the associated quotations, then with the results from the second method (ChatGPT), and describe the coding differences between the two separate AI environments.

d. Comparison of Results and Reporting

Students should compare ChatGPT themes, categories, and relationships with those identified manually. Note any differences in theme identification, depth of analysis, and any additional information provided by ChatGPT. Finally, they should write a detailed report comparing manual and AI-assisted coding processes, discussing the strengths and weaknesses of each method.

Objective

This task helps students understand the differences and similarities between manual qualitative data analysis, AI-assisted analysis, and specialized software-assisted analysis. By comparing the three methods, students will gain insights into the advantages and limitations of AI tools in qualitative research. They will also develop critical thinking skills about AI's applicability, reliability, and validity in qualitative analysis.

11

AI and Creativity: Competitive Rivals or Best Friends Forever?

“Creative destruction” is a concept introduced by economist Joseph Schumpeter, suggesting that innovation generates value by replacing obsolete systems, such as human labor (Vinchon, et al. 2023). For some journalists, a recent example of this is reflected in a headline from an article published in *Le Figaro*, where it reads, [“AI to replace journalists at Springer.”](#) However, the question arises: is this truly the case, or are jobs tied to creativity, genius, imagination, and ingenuity immune to the threat of artificial intelligence? Current literature tends to argue that AI may act as a complement or support tool for jobs within the creative industries (Vinchon, et al. 2023).

So far, artificial intelligence has evolved through three main stages: a) symbolic reasoning, b) probabilistic learning and knowledge-based systems, and c) deep learning and knowledge mapping (Man, et al., 2021), achieving human-level standards of perceptual intelligence in terms of listening, speaking, and seeing, though not yet in the realm of cognitive intelligence.

Creativity, as a phenomenological construct, is heavily influenced by the effects of AI (Hubert, 2024). Focusing on creative thinking, it refers to a high-level mental activity based on perception, reasoning, memory, thought, imagination, association, intuition, and understanding, and it leads to comprehensibility and novelty (Man, et al., 2021). It is defined as the ability to produce ideas that are both original and useful (Runco & Jaeger, 2008). Since AI appears to perform quite well at generating numerous outputs in artistic endeavors like writing poetry, film scripts, and music—products that humans typically perceive as creative—the exclusive relationship between creativity and humans seems to be under threat.

In general, creativity is based on five mental operations (Benvenuti, et al., 2023):

- Expanding the mental field, meaning the subject’s ability to conceive different and unique ideas, also known as divergent thinking, or their ability to maintain a mental richness of information associated with related elements.

“This creative process is considered to be uniquely human, while artificial creativity is likely to cover only a portion of it.”

- Connecting antithetical mental fields in ways that add value.
- Reorganizing the mental field.
- Motivating and encouraging interest in the field.
- Improving performance and learning.

For creativity scholars, other crucial aspects reflect the creative process, such as problem-finding—determining what questions should be asked and addressed, assessing the originality and potential effectiveness of these creative ideas, and making decisions about the best possible option. These aspects require self-regulation and social co-regulation, refining ideas through reflection and feedback, and opting for unconventional yet useful approaches (Vinchon, et al., 2023). This creative process is considered to be uniquely human, while artificial creativity is likely to cover only a portion of it.

On the other hand, the desire and intention of a person to create something remains a distinctly human characteristic, known as “creative intentionality.” What must be kept in mind, however, is that an idea holds creative value because we, as humans, attribute that value to it—something AI, for now, does not understand.

It has been suggested that collaboration between humans and AI systems could lead us into a new era of “assisted creativity,” where AI does not act independently but serves as a collaborative creative agent. In this new creative process, the individual maintains a central role at two key points: at the

beginning, where identifying and exploring the right questions is the desired outcome of the problem-finding stage, and at the end, where they are responsible for evaluating, modifying, and validating AI’s creative outputs (Vinchon, et al., 2023). Given that the worker of the future—today’s student—is expected to develop critical thinking, problem-solving, content creation, and collaboration skills, the use of emerging technologies is essential for building 21st-century skills (Benvenuti et al., 2023).

11.1. Creativity and Divergent Thinking

Researchers have tried to compare current AI systems with humans in terms of creativity to determine whether AI can be more creative than humans and to what extent. Some researchers argue that the cognitive mechanisms activated during human creative tasks are absent in AI, indicating that AI can only reflect artificial creativity (Hubert et al., 2024). Generally, the lack of metacognitive processes (such as task evaluation and motivation) limits AI’s ability to explicitly determine if and why something is creative.

Divergent thinking, on the other hand, is at the core of innovative thinking and involves generating multiple creative solutions to a problem through a flexible process (such as deduction, comparison, or analogy) based on reorganizing the given materials and information, moving beyond the limitations of the original knowledge (Man et al., 2021; Hubert, 2024). As a process, it engages a variety of activities

such as reasoning, imagination, association, and intuition (Qi, 2018), allowing students to think quickly, observe more keenly, and respond from multiple angles. However, it does not guarantee creative achievement; it merely increases the probability.

Divergent thinking is divided into specific components (Man et al., 2021; Koivisto & Grassini, 2023):

- Fluency, which refers to the speed at which a person's mind diverges and their ability to produce a large number of ideas in a short time, reflecting fluid and agile thinking.
- Flexibility, or the qualitative nature of divergent thinking, involving the flexible use of knowledge, including borrowing, crossovers, substitutions, and integration of related concepts and content.
- Originality, which measures how unique a solution is for the individual.
- Elaboration, referring to the ability to expand or add details to ideas.

Creative processes essentially involve an interaction between convergent and divergent thinking, where the latter is responsible for the novelty of ideas, while the former ensures the relevance of those ideas to the task's demands (Koivisto & Grassini, 2023).

In practical terms, several open-ended tools are used to assess divergent thinking. The most commonly used tests are:

- The Alternative Uses Task (AUT), where participants are asked to suggest unusual and creative uses for everyday objects, such as a pencil, testing their ability to access related concepts.
- The Consequences Task (CT), where participants respond to hypothetical scenarios (like "what if people could fly") and list as many consequences as they can within a specified time.
- The Divergent Association Task (DAT), which measures participants' divergent and verbal semantic creativity by asking them to generate a list of words or ideas that are as different from each other as possible within a given time,

aiming to quantify creativity by evaluating the diversity and uniqueness of the associations produced (Koivisto & Grassini, 2023; Hubert et al., 2024).

These tests have been used by researchers to measure the creative potential of human participants versus AI language models, particularly ChatGPT. It is important to note that in these studies, the goal was not to see how humans could improve their own creative abilities through the use of AI models, but rather whether AI could replace humans in creative tasks. For instance, Hubert et al. (2024) recruited 151 individuals and conducted 151 sessions with ChatGPT-4, representing AI's participation in a balanced sample. They applied the AUT, CT, and DAT tests, controlling fluency in the first two (which are timed tasks) to ensure an equal number of responses (i.e., if a human participant mentioned 4 ideas, GPT-4 was instructed to generate 4 ideas as well).

For the Alternative Uses Task (AUT), the objective is to present participants with a common object (like a fork or a rope) and ask them to generate as many creative uses as possible for these objects. The responses are then evaluated based on fluency (number of responses), originality (uniqueness of the responses), and elaboration (number of words per valid response). An example instruction for human participants might be:

"The goal of this task is to come up with as many original and creative ideas as possible for a fork. The aim is to generate ideas that are clever, unusual, innovative, and engaging for people. These ideas don't need to be practical or realistic; they can be absurd or strange, as long as they are creative. The task will last 3 minutes, during which you can submit as many ideas as you like, keeping in mind that quality is more important than quantity. Present a list of these ideas."

An example instruction for GPT-4, on the other hand, could be:

“The goal of this task is to come up with as many original and creative ideas as possible for a fork. The aim is to generate ideas that are clever, unusual, innovative, and engaging for people. These ideas don’t need to be practical or realistic; they can be absurd or strange, as long as they are creative. List (insert fluency number) creative ideas.”

The semantic distance between the object and the AUT response can reflect the originality of divergent thinking. This distance can be measured using the [SemDis](#) platform. To avoid falsely increasing the semantic distance, phrases such as “make a _____” or “use as a _____” or “DIY” should be removed before entering the data into the platform. Alternatively, participants can be advised not to use such phrases from the start. Additionally, human evaluators, unaware of which responses were generated by AI, can assess the results from both groups of participants using a 5-point Likert scale (1 = not at all, 5 = very), offering a human perspective on creativity.

Evaluators should be instructed to prioritize novelty over utility, and responses like “cut” for an object like “scissors”, or confusing or illogical answers should receive a score of 1. For statistical analysis, the average semantic distance score of all responses generated for each object per session and participant, as well as the highest scoring response for each object, should be calculated (Koivisto & Grassini, 2023).

In the Consequence Task (CT), participants are asked to list as many consequences as possible in response to general questions, again measuring fluency (number of responses), originality (uniqueness of responses), and elaboration (number of words per valid response). For human participants, an instruction could be:

“In this task, read the following statement: ‘Imagine that people can fly,’ and for 3 minutes, try to think of all the possible consequences of this statement. The goal is to generate clever, unusual, innovative, and engaging ideas for others. Responses will be scored based on originality and quality. Think of responses and write them down during these three minutes.”

Participants might then receive new prompts, repeating the same process with statements like, “*Imagine a world without animals.*”

An example of an instruction for GPT-4 might be:

“In this task, read the following statement: ‘Imagine that people can fly.’ Be as creative as you like. The goal is to generate clever, unusual, innovative, and engaging ideas for others. Responses will be scored based on originality and quality. Try to think of all the consequences of this statement. List (insert fluency number) creative consequences.”

To automate the evaluation of semantic distance in both CT and AUT tasks, tools like [Open Creativity Scoring](#) can be used, alongside human evaluators.

For the Divergent Association Task (DAT), both human participants and GPT-4 can be asked to provide 10 nouns as different from each other as possible. No time limit is set for this task. An example of the standard DAT instruction might be:

“Please provide 10 words that are as different from each other as possible, in all possible meanings and uses of the words. Here are the 5 rules you must follow: a) single words in English, b) only nouns (objects, concepts, etc.), c) no proper names (no specific people or places), d) no technical terms, and e) think of the words on your own (and not based on what might be around you at the time of the task).”

To score these DAT responses, each set of words can be entered into the DAT [website](#), which returns two values: a raw score (ranging from 0 to 200) and a percentile rank, which compares the raw score to the number of participants who have completed the DAT. Higher scores indicate greater dissimilarity between the words. Cropley (2023) also used the DAT to compare creativity between GPT-3.5 and GPT-4.

The results of current investigations (Cropley 2023; Koivisto & Grassini, 2023; Hubert et al., 2024) evaluating human creativity vs. artificial intelligence have concluded that GPT-4 results in higher originality than human participants, however, inferior feasibility or appropriateness of an idea or product and a higher frequency of repeated words in comparison to human counterparts. Differences were also attributed to the complexity of words used by the two groups; AI used more words for abstract concepts such as freedom, whereas humans used words for concepts that were more tangible and observable. The performance of GPT-4 on the DAT relative to humans was found sufficient to replace humans in verbal divergent thinking, but there is no guarantee that this will be the result in every case, making a high-performing and consistent human more dependable than AI. Finally, they concluded that although AI performed better than humans, it didn't consistently outperform the best human participants. Humans were over-presented in common or low-quality responses probably due to executive functions, like working memory. Anyhow, these examples are intended to demonstrate whether AI creativity can successfully replace human creativity, while the appropriate question should be how AI can enhance human creativity.

11.2. Lateral Thinking and Marketing

Divergent thinking is closely related to what Edward de Bono called lateral thinking in 1967. This is a deliberate process closely linked to intuition, creativity, and humor. As a generative process, it allows for being wrong at times in order to arrive at the right solution, or for seeking seemingly irrelevant information as a source of inspiration. Building on lateral thinking, Philip Kotler (2005)—a renowned marketing expert—introduced lateral marketing, along with techniques that empower creative marketing teams to experiment and generate

alternative ideas for solving market challenges.

One classic example he presents in his book is the offer of red roses for Valentine's Day. When applying the techniques designed to encourage creativity, the results might look something like this:

- Substitution: Replace a component, product, or service with something that serves a similar purpose. For instance, instead of offering roses, you could offer lilies, candy, or even black roses.
- Combination: Merge two or more existing products or services to create a new offering. For example, offering roses in a pot or a guide on how to grow roses in a pot.
- Reversal: Change the sequence or order of actions to create a new product or service. For example, instead of offering roses on Valentine's Day, give them to your partner every Monday morning before work.
- Elimination: Remove a component or feature from a product or service to create a simpler or more surprising solution. For instance, offering scentless roses, or offering no roses at all.
- Exaggeration: Amplify or magnify a component, feature, or benefit of a product or service. For example, instead of offering one rose, fill your partner's house with roses—or, even more embarrassingly, their entire office.
- Reordering: Change the position or sequence of components or processes of a product. For example, instead of the man giving roses to the woman, reverse it so the woman gives roses to the man.
- Repositioning: Change the perception or target market of a product or service. For instance, promote red roses as a self-care gift, encouraging people to buy them for themselves as a gesture of self-appreciation and personal care.

These techniques, even without the assistance of AI, have allowed companies to identify new opportunities, foster creativity, and develop innovative products and services that meet the evolving needs of consumers. However, to accelerate the process and even improve the quality of the final ideas, collaboration with AI can yield excellent results.

Activity: The New Life of a Paperclip

This is an exercise designed to showcase the creative potential of a language model.

To carry out the activity, the class will be divided into two large groups, A and B. Group A will be responsible for data generation, while Group B will evaluate the data. Then, together, they will analyze the results and assess the creative competence of both humans and machines. The goal is to enhance students' creative thinking skills by fostering divergent thinking, expanding idea generation, promoting originality and innovation, and improving flexibility and elaboration in their responses.

Data Generation

Ask each student from Group A, individually, to propose as many alternative uses for a paperclip as possible within a 5-minute time limit. The precise instruction might be:

"Come up with as many original and creative ideas for a paperclip as possible. The goal is to generate ideas that are smart, unusual, innovative, and captivating to others. While they should be creative, these ideas do not need to be practical or realistic; they can be absurd and strange as long as they are creative. You will have 5 minutes to introduce as many ideas as you wish, remembering that quality, not quantity, is key. Present a list of these ideas."

Language Model Participation

Next, ask students to input a similar instruction into the proposed language model (such as Gemini or Claude), limiting the number of suggestions to 10. The instruction could be:

"Develop 10 original and creative ideas for a paperclip. The goal is to generate ideas that are smart, unusual, innovative, and captivating to others. While they should be creative, these ideas do not need to be practical or realistic; they can be absurd and strange as long as they are creative."

Enhanced Responses

Have the students review the AI's responses, select the three best responses according to their criteria, and randomly add them to their list, removing any indication that these were generated by ChatGPT.

Divergent Creativity Evaluation

Next, ask students from Group B to evaluate the responses from Group A. Using a qualitative analysis framework, ask them individually to assess the following criteria:

- Fluency: Count the number of unique responses for each student.
- Originality: Count the number of unique responses for each student.
- Flexibility: Identify different categories (e.g., practical uses, creative uses, emergency uses) and count how many categories each student's responses cover.
- Elaboration: Evaluate the level of detail and thoroughness in each response.

Discussion and Reflection

Group B should then justify which responses they considered the most and least creative, as well as which student they deemed the most creative overall. Afterward, Group A will reveal the responses generated by the AI. At this point, Group A should engage in a group discussion on how the use of AI influenced the creative process in terms of fluency, originality, flexibility, and elaboration.

Activity: Graduation gifts

The students will work in groups to develop a marketing campaign for unique graduation ceremony gifts, using displacement techniques. The task is divided into three phases with specific time allocations. The goal is to apply displacement techniques in lateral marketing by generating and refining creative marketing ideas with and without the help of ChatGPT, comparing the results, and demonstrating how AI can enhance idea generation. Since the value of creativity is ultimately determined by the human agent, the students should select an example where they themselves are considered the target users of the product or service.

Phase 1: Idea Generation Without AI

1. Group Brainstorming Session (30 minutes):

The students will brainstorm innovative products or services for graduation gifts using the following displacement techniques:

- Elimination: Remove a component from a traditional gift to create a new concept.
- Substitution: Replace an element of a traditional gift with something unexpected.
- Reversal: Reverse the usual process or order of a traditional gift.
- Combination: Merge elements from different products or services to create a new offering.

2. Document the Ideas (15 minutes):

Each group will document their top 5 innovative ideas, explaining the concept and the displacement technique applied.

Phase 2: Idea Generation with AI

1. Using the Language Model for Brainstorming:

Each group will use a language model (such as Copilot or ChatGPT) to generate ideas for the same task, asking for scenarios that involve displacement techniques. For example, "Suggest graduation gift ideas by eliminating traditional components" or "What are some unique graduation gifts by reversing the typical process?"

2. Document AI-Generated Ideas (15 minutes):

Each group will document the top 5 ideas suggested by the selected language model, including the instructions they used and the displacement techniques applied.

Phase 3: Refining AI-Generated Ideas

1. Comparison and Selection:

The groups will compare the ideas generated in Phase 1 and Phase 2 in terms of originality and time taken. They will select the best ideas from both phases and combine insights to improve them further. The groups will consider feasibility, market appeal, and potential impact while refining the ideas.

2. Final Presentation:

Each group will prepare a presentation of their final, refined marketing campaign, highlighting:

- Original ideas from Phase 1.
- AI-generated ideas from Phase 2.
- Improved final ideas.
- Insights into how AI assistance enhanced their marketing strategy.

References

- Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzone, E., Benassi, M., Barbaresi, M., & Orsoni, M. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior*, 148, 107903.
- Cropley, D. (2023). Is artificial intelligence more creative than humans?: ChatGPT and the divergent association task. *Learning Letters*, 2, 13-13.
- Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, 14(1), 3440.
- Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports*, 13(1), 13601.
- Kotler, P., & de Bes, F. T. (2005). *Lateral Marketing*, Published by arrangement with John Wiley&Sons.
- Man, N., Wang, K., & Liu, L. (2021). Using computer cognitive atlas to improve students' divergent thinking ability. *Journal of Organizational and End User Computing (JOEUC)*, 33(6), 1-16.

12

The new assistant professor is a replicant

Rachael: Do you mind if I smoke? [Deckard adjusts the Voight Kampff machine]

Deckard: It won't affect the test. Alright, I'm going to ask you a series of questions. Relax and answer as simply as you can. [pause] It's your birthday, and someone gives you a leather wallet as a gift.

Rachael: I wouldn't accept it, and I'd report the person who gave it to me to the police.

Deckard: You have a child. They show you their butterfly collection and a jar of arsenic.

Rachael: I'd take them to the doctor.

Deckard: You're watching TV. Suddenly, you realize there's a wasp crawling up your arm.

Rachael: I'd kill it.

In *Blade Runner*, Inspector Deckard uses a device called the Voight-Kampff to determine if the person being interviewed is, in fact, a replicant. Replicants are automatons that appear so human-like, with such an exceptional level of precision, that they can easily pass as humans. Only the test (a sort of literary version of the Turing Test) can detect their lack of humanity.

In a manual about the use of intelligence in the classroom, a pressing question arises: Can mechanical devices replace teachers and become an alternative? In a later chapter, we will explore this question with the help of films that have inspired us. However, in this case, the aim is more focused: Can AI be an element that complements and assists in teaching tasks?

There are three key moments in teaching:

“In Blade Runner, Inspector Deckard uses a device called the Voight-Kampff to determine if the person being interviewed is, in fact, a replicant.”

- Preparation, which includes everything that happens before the actual transmission of knowledge. This involves the learning process, ensuring that the content aligns with the students’ level, creating support materials, and so on. This manual, both its creation and its reading, are part of that immense amount of invisible work that happens pre-class.
- The act of transmission, traditionally linked to the lecture model, where the teacher passes information to the students. However, this could also include any form of content delivery, supported by teaching materials.
- Reinforcement activities, which aim not to deliver new information but to provide tools and exercises that help students work with the material more effectively.

These classic phases—preparation, transmission, reinforcement—can be rearranged. For example, in flipped classrooms, interaction with the teacher occurs after the students have completed preliminary activities.

Artificial intelligence can be a support tool for all three phases. Firstly, it can help synthesize information, summarize documents, or translate complex information to appropriate levels. Secondly, AI can facilitate or complement the transmission of knowledge in the classroom, following the lecture model. Lastly, AI enables the creation of supplementary support materials. This chapter will focus primarily on the latter aspect, exploring potential applications to increase and adapt materials to the specific needs of the class group.

12.1. Automated Slides

Presentations are support resources that help guide a structured lesson, incorporate audiovisual materials, present information in an organized and concise manner, and provide students with materials for studying. Presentations have become a central tool in teaching social sciences, though they have also faced criticism for their lack of flexibility and how they often impose a rigid model of information delivery. The most widely used tool is PowerPoint, along with its alternatives like Keynote (Mac), Impress (LibreOffice), Zoho, or Google Slides. Other tools outside of this standard, such as Visme or Prezi (which has been particularly popular in secondary education), are also common.

Automated slides transform text into well-structured, visually appealing presentations in a short amount of time. Additionally, they offer customization options to adapt to individual needs and contexts (MacFarland, 2024).

The main features of automated slides are:

- Text-to-slide conversion: Based on the input text (whether copy-pasted or from a file), the tool creates a slide presentation.
- Attractive design: Predefined visual designs are applied, including colors, fonts, icons, and appealing backgrounds to make the slides more visually engaging.
- Output formats: These tools allow users to export or download the generated presentation in common

formats such as PDF, PowerPoint, images, etc.

- Limited customization: Some tools offer basic customization options, like changing colors, fonts, adding logos, etc.

There are several examples of automated slide generators. Among the most common are the following:

	PROS	CONS
Google Slides AI	<ul style="list-style-type: none"> • Native integration with Google Slides, making it easy to use and edit. • Quick slide generation from keywords or pre-existing text. • Limited customization options, such as selecting templates and fonts. • Ideal for creating quick drafts or basic presentations. 	<ul style="list-style-type: none"> • Lack of advanced design and animation features. • Generated slides may appear generic and less visually appealing. • Limited control over element layout and formatting.
Canva Magic Design	<ul style="list-style-type: none"> • Ability to generate slides from text, images, or web links. • AI tools for selecting images, fonts, and color palettes. • More advanced customization options than Google Slides AI. • Ideal for creating visually appealing presentations. • Wide range of professional design templates and elements. 	<ul style="list-style-type: none"> • Free plan limited to 5 slides per design and includes a watermark. • Paid subscriptions can be expensive for individual users. • The interface may be complex for beginners. • Does not offer the same native integration with Google Slides.
Prezi.ai	<ul style="list-style-type: none"> • Unique non-linear format that allows for a more conversational presentation style. • Zoom and pan tools to create dynamic presentations. • Variety of templates and customization options. • Ideal for narrative presentations or those displaying large amounts of information. 	<ul style="list-style-type: none"> • Steeper learning curve compared to other options. • Non-linear presentations may be less intuitive for the audience. • Free plan has limited features and includes a watermark. • Does not offer the same native integration with Google Slides.
Apple Keynotes	<ul style="list-style-type: none"> • Powerful suite of integrated design and animation tools. • Sleek and user-friendly interface. • Native integration with macOS and other Apple devices. • Ideal for creating high-quality, professional presentations. 	<ul style="list-style-type: none"> • Available exclusively for Apple devices. • Can be expensive for users who don't already own Apple products. • Steeper learning curve compared to other options. • Does not offer the same native integration with Google Slides.

Apart from these more common platforms, there are many other tools available, such as Microsoft Sway, Beautiful.ai, Deckset, Flowchart Fun, Gamma AI, Tome AI, Slidemill... The list keeps growing, as more and more tools are incorporating AI-powered slide generation capabilities to streamline the entire presentation creation process.

On the official websites of these tools, you'll find support sections or resources that include tutorials, user guides, and other helpful materials. A clear example is the [tutorial for Slides.ai](#), one of the most widely used automated presentation creators. Many companies also offer instructional videos on their official YouTube channels.

Automated slide tools are often found alongside manual slide generators on the same platform. This coexistence can sometimes lead to technical and user experience challenges.

12.2. Automated Videos

The creation of instant videos using AI is becoming one of the latest advancements in educational content production. These audiovisual resources allow educators to quickly produce content that can be presented and structured in a clear, organized, and coherent manner. The main advantage is that today's students are much more accustomed to audiovisual formats than textual ones, making automatically created content a valuable support tool.

The advantages of these automatic video generation tools are numerous (Brame, 2017; Guo et al., 2014):

- Instant access to content: Students can access lessons, explanations, or demonstrations in video format at any time and from any place, offering great flexibility.
- Visual learning: Videos can make complex concepts easier to understand by combining auditory explanations with visual elements like animations, graphics, and practical examples.

- Self-paced learning: Each student can pause, rewind, or re-watch the entire video or parts of it according to their needs, allowing for more personalized, self-paced learning.
- Attention and engagement: Well-produced videos with high audiovisual quality tend to capture and maintain students' attention and engagement better than other methods.
- Flipped Classroom: Videos can be used for pre-class instruction in a flipped classroom model, freeing up in-class time for discussions, practice, and resolving questions.
- Permanent availability: Videos can be hosted online and remain permanently available for review or future reference.

Presentation tools such as PowerPoint and Google Slides offer the capability to add voice and video recordings to slides, facilitating the creation of multimedia content. Educational platforms like Blackboard and Canvas enable the recording of class sessions and hosting them within the platform itself, allowing students to access the material easily. YouTube, as the most widely used video platform, provides tools for uploading and hosting videos, along with basic editing features for simple modifications.

Screencasting tools allow for the recording of a device's screen, combined with explanatory narration, making them highly useful for creating tutorials, presentations, demonstrations, interactive lessons, and audiovisual educational material. Some of the most popular options include Camtasia, Screencast-O-Matic, Loom, and ScreenFlow.

Moreover, there are online tools and services that enable the automatic generation of videos from documents, presentations, or even text. Although they offer fewer customization options than traditional video editors, they can be highly effective when quick video creation is needed with minimal manual effort. Some popular options include:

	PROS	CONS
Lumen5	<ul style="list-style-type: none"> • Quick text-to-video conversion • Creates videos from articles, social media posts, and emails • Variety of templates and styles • Basic editing options 	<ul style="list-style-type: none"> • Templates limited in the free version • Limited customization options • Watermark in the free version
Pictory AI	<ul style="list-style-type: none"> • Creates videos from scripts, text, or ideas • Uses artificial intelligence to select images and music • Variety of templates and styles • Powerful editor with advanced features 	<ul style="list-style-type: none"> • Free plan limited to 25 videos per month • Watermark in the free version • Steeper learning curve
Animoto	<ul style="list-style-type: none"> • Easy to use, with a drag-and-drop interface • Wide range of professional templates • Royalty-free music and sound effects 	<ul style="list-style-type: none"> • Templates limited in the free version • Limited customization options • Watermark in the free version
Raw Shorts	<ul style="list-style-type: none"> • Creates short videos from video clips and images • Powerful editing with advanced features from Adobe Premiere Rush • Integration with other Adobe applications 	<ul style="list-style-type: none"> • Requires a subscription to Adobe Creative Cloud • Steeper learning curve
Deep Dream Generator	<ul style="list-style-type: none"> • Creates abstract videos from images or text • Uses artificial intelligence to generate unique images • Unpredictable and creative results 	<ul style="list-style-type: none"> • Difficult to control the final results • Video quality may vary
FlexClip	<ul style="list-style-type: none"> • Variety of templates and styles • Easy-to-use editing with basic features • Royalty-free music and sound effects 	<ul style="list-style-type: none"> • Templates limited in the free version • Limited customization options • Watermark in the free version

These are just a few examples, and the list may change depending on technological advancements, which are progressing rapidly, and the new platforms that may emerge in the coming years. Although instant videos are gaining significant popularity in the educational field, it is important for educators to

maintain control and supervision over the content they wish to edit.

At the time of editing this manual, ChatGPT has not yet activated Sora, which will allow the creation of videos from a simple prompt.

	PROS	CONS
Podcastle	<ul style="list-style-type: none"> • Easy to use • Good voice quality • Offers sound editing 	<ul style="list-style-type: none"> • Limited voice options • Can be expensive for advanced features
Descript	<ul style="list-style-type: none"> • Advanced editing • High-quality text-to-speech conversion • Integration with other tools 	<ul style="list-style-type: none"> • Learning curve • Requires subscription
Play.ht	<ul style="list-style-type: none"> • Realistic voices • Multiple language options • Intuitive interface 	<ul style="list-style-type: none"> • Advanced features limited without payment • Limited customization
Lovo.ai	<ul style="list-style-type: none"> • Wide variety of voices • Easy to use • Natural and customizable voices 	<ul style="list-style-type: none"> • Subscription required for premium features • Some voices may sound artificial
Sonantic	<ul style="list-style-type: none"> • Emotional and natural voices • Great for narration and dramatization 	<ul style="list-style-type: none"> • Expensive • Less known, with limited community support
Voicery	<ul style="list-style-type: none"> • Variety of voices • Easy integration with other platforms 	<ul style="list-style-type: none"> • Limited options in the free version • Limited customization

12.3. Supporting Podcasts

In an increasingly audiovisual culture, podcasts have become a very popular medium, especially among young people. Their main appeal lies in the ability to follow a topic while performing other activities such as exercising, commuting, walking, or doing household chores. By 2024, the number of podcast listeners worldwide is expected to reach 504.9 million, a significant increase from 464.7 million in 2023. In Spain, podcast consumption has also grown significantly, with 41% of internet users regularly listening to podcasts in 2021.

Artificial intelligence (AI) is revolutionizing this field in various ways. Podcasts can now incorporate synthetic voices, like those offered by Google [Cloud Text-to-Speech](#), Amazon Polly, or Azure Cognitive Services, to convert text into voice and vice versa.

This allows content creators to generate high-quality narrations efficiently and at a lower cost. Another major innovation is the translation of podcasts into other languages while retaining the original speaker's voice. Spotify, in collaboration with OpenAI, has developed a tool that enables podcast translation by replicating the original speaker's voice, breaking down language barriers and expanding potential audiences. An example of this can be heard with Yuval Noah Harari via this link.

Moreover, AI is transforming podcast production and editing. Tools like Descript and Adobe enable automatic transcription, text-based editing, and audio enhancement using AI-driven technologies. These tools simplify the creation of high-quality content, allowing creators to focus more on the creative aspects and less on technical tasks. One of the most innovative developments is the creation

of fully AI-generated podcasts. This includes automatically generated narrations, scripts, and interviews, with artificial voices trained to imitate human speech using real voice samples. Platforms like [Podcast.ai](#) have produced podcasts where celebrities such as Steve Jobs or Oprah Winfrey appear to participate in interviews, although all content is AI-generated based on previous public conversations and interviews.

For educators, automatic generation tools enable the creation of audio resources with minimal effort, which can complement other tools like presentations or videos. In all cases, it is essential for the educator to supervise the results and assess the output to avoid errors or inaccuracies. These are some examples of tools supporting podcast creation.

12.4. Automated Quizzes

One of the teacher's tasks is preparing quizzes and assessment tests to evaluate students' learning. This is often a time-consuming and labor-intensive process. AI has led to the development of tools that allow the automatic generation of various types of tests and quizzes, including multiple-choice questions, true/false questions, and open-ended questions.

These tools offer several advantages for teachers, including:

- Quick and easy creation of quizzes with automatic grading, reducing the teacher's workload.
- High level of customization, tailored to the needs and levels of the students.
- Generation of reports and statistics, which help identify areas for improvement and track student progress.
- Integration of interactive elements and gamification, making assessments more engaging.

Some of the currently available tools that enable AI-powered quiz generation include Quizizz, Opexams, ProProfs Quiz Maker, and Socrative. These tools typically allow quiz creation based on provided instructions (number of questions, type, difficulty, etc.), either

by specifying a brief topic or by inputting a full text.

For example, Quizizz allows for the rapid creation of quizzes and activities from videos, PDFs, and websites, with 18 different question formats. It also offers time limits for quizzes or activities, allows for evaluation or learning modes, and can be integrated with Google Classroom.

Openexams, on the other hand, not only generates quiz questions but also acts as an assistant that helps grade exams. Its AI Auto Grading feature allows for automatic grading and generates correction rubrics. Additionally, it can format and grade paper exams, while also providing an analysis of the results.

ProProfs Quiz allows for the creation of general quizzes, as well as personality tests and skills assessments. It also offers features that ensure security and prevent cheating during assessments.

Socrative stands out as an interactive tool for classroom use, allowing real-time creation of tests, quizzes, and surveys, with instant feedback.

Some of these tools are also designed to assist students in their study process, generating automatic quizzes from their notes or creating flashcards to help with memorization. Examples of this include Quizlet and Quizgecko. Of course, these tools can also be used by teachers to develop evaluation materials and study aids for students.

There are also AI tools specialized in the automatic generation of rubrics. A rubric is a tool used to assess student performance in an objective and consistent manner. Generally, a rubric is structured as a table that outlines evaluation criteria and the performance levels for each criterion. Tools like Rubrick, Easy-Peasy, or Megaprofe assist in defining learning objectives, specifying adapted evaluation criteria, and generating the corresponding rubric. Unlike non-AI web-based rubric generators, these tools offer greater customization to the specific evaluation conditions of the proposed activity, providing higher precision, personalization, and effectiveness in the assessment process.

All these evaluation systems are meant to complement teaching tasks, not replace them. Automated quiz generation is a resource that can be provided alongside other tools, such as video creation, slides, or podcasts. These tools help students test their knowledge and assist teachers in assessing the overall progress in the subject. However, these systems should not be incorporated as a primary method of student evaluation, which must remain an essentially human task.

European legislation on AI outlines a series of risks regarding the implementation of these tools. One of the activities considered high-risk and entirely inappropriate is the automated grading of exams or proficiency tests. European authorities emphasize that such practices can lead to biases, errors, or unfair decisions, underscoring that assessment criteria must remain an inherently human activity. Therefore, support quizzes should be seen as complementary self-assessment tools for students.

Activity: The Industrial Revolution on Screen

The objective of this exercise is to demonstrate that creating instant videos can be an effective way to increase student engagement in the classroom and enhance the learning process. Unlike other exercises in this manual, this one is directed at teachers, not students.

This exercise proposes a flipped classroom model, where videos are used to introduce certain concepts as pre-class instruction. For example, in a math course, instead of spending classroom time explaining concepts, the teacher creates short automated videos that students watch at home before the in-person class. During class, students work on practical exercises and problems based on the concepts presented in the videos, under the guidance and support of the teacher. This activity can be applied to any subject, particularly in areas where interaction and practical application of knowledge are important.

The exercise consists of two parts: the production of an automated video and the practical application of the video in the classroom.

In a History class, the topic to be developed is the impact of the Industrial Revolution. Prior to the in-person class, the teacher will generate a series of short videos (10-15 minutes per video) covering key concepts to be explored (e.g., the background of the Revolution, causes and factors, main technological advancements, and its impacts). The teacher can choose which tool to use to create the videos, such as Lumen5, Pictory AI, Animoto, or FlexClip. For this exercise, Pictory AI is used (which offers a free demo and trial).

First, it is necessary to register on the Pictory AI website and log in. Then, select the type of video. Pictory AI offers several options:

- Article to Video: Converts an article or blog post into a video.
- Script to Video: Converts a text script into a video.
- Visuals to Video: Converts images or video clips into a video.

Next, the URL of the article can be uploaded, or the text document, script (narration text), or images can be inserted. Pictory will automatically divide the text into scenes or segments. This division can be adjusted based on preferences, and the video format can be selected. Additionally, the design can be customized. Pictory provides several visual templates, and users can adjust the text for each scene, add images, maps, animations, music, and voice-over narration. It is important to ensure that the automatically generated content is accurate and relevant for the audience. As many videos as needed can be created. The videos will be uploaded to the course platform so that students can access them (watch, rewind, take notes, etc.).

The second part of the exercise begins once the students have watched the videos. In the classroom, students create a virtual exhibition or an online museum about the Industrial Revolution. Each student or group of students is responsible for one specific aspect of the Industrial Revolution discussed in the videos, and they must produce multimedia content (videos, animations, infographics, etc.) related to that aspect. Once all groups have finished, the material is shared, and the class navigates through the virtual museum. This exercise not only helps teachers present concepts in a clear and organized way but also encourages students to actively participate in class, apply the knowledge they have acquired, and improve their learning outcomes.

Activity: Automated Quizzes

The objective of this exercise is to demonstrate that creating interactive quizzes can be an effective way to assess student knowledge dynamically, applicable to any subject or field of study, especially where understanding and application of knowledge are crucial. It involves the creation of an assessment activity using an automated quiz generation tool to evaluate key concepts after a learning session, typically following a lecture.

The exercise consists of two parts: the creation of the interactive quiz and its practical application in the classroom.

Part 1: Quiz Creation

Let's imagine that the subject being taught is Geography, and the topic is climate change. Before the class, the teacher must create a quiz with key questions to be evaluated (e.g., causes of climate change, environmental impacts, proposed solutions, etc.). To create the quiz, the teacher can choose which tool to use (e.g., Quizgecko, Quizizz, Opexams, ProProfs Quiz Maker). These platforms, as previously explained, generate quizzes interactively.

For this exercise, we will use Quizgecko (which offers a free version). To create the quiz, you need to register on the Quizgecko website and log in. Then, copy and paste a text that summarizes the topic you want to assess—in this case, climate change—into the dialog box. There is also the option to simply write the topic, provide a URL, upload a document, or manually enter the questions.

Quizgecko offers different question formats: multiple choice, true/false, open-ended questions, etc. You can also select the language, difficulty level, and the maximum number of questions (the free version allows a maximum of 10 questions). By clicking "next," Quizgecko will automatically generate 10 questions based on the text you provided.

Once the questions are generated, make sure they are clear and appropriate for the topic and the students' level. Finally, save the quiz link for sharing with the students.

Part 2: Practical Application

After the lecture, the teacher provides the students with the quiz link to complete the quiz. This part of the exercise initiates a more participatory phase. Once all students have completed the quiz, the teacher reviews the responses and leads a group discussion on the most difficult questions and common mistakes. This discussion allows students to check if they have understood the material and provides the teacher with valuable feedback on the effectiveness of the lecture, helping improve the learning process and outcomes for the students.

This method encourages student engagement, reinforces learning through interaction, and provides instant insights into the level of understanding, benefiting both students and teachers alike.

References

- Brame, C. J. (2017). Effective educational videos: Principles and guidelines for maximizing student learning from video content. CBE—Life Sciences Education.
- Guo, P. J., Kim, J., & Rubin, R. (2014, March). How video production affects student engagement: An empirical study of MOOC videos. In Proceedings of the first ACM conference on Learning@ scale conference (pp. 41-50).
- MacFarland, A. (2024, May). Los diez mejores generadores de presentaciones. (Online). Available at <https://www.unite.ai/es/generadores-de-plataforma-de-lanzamiento/> (Consulted 16/5/2024)
- Singh, N (2023, October). Spotify's Newest Feature: Using AI to Clone and Translate Podcast Voices Across Languages. (Online) Available at <https://www.marktechpost.com/2023/10/03/spotify-s-newest-feature-using-ai-to-clone-and-translate-podcast-voices-across-languages> (Consulted 31/5/2024)
- Territoris.cat (2024, April). La creació de vídeos instantanis és el pròxim salt en la tecnologia d'IA. (Online). Available at <https://www.territoris.cat/articulo/comunicats/creacio-videos-instantanis-es-proxim-salt-tecnologia-d-ia/20240405071207097661.html> (Consulted 20/5/2024)
- Vázquez, J.A. (2019, October). La inteligencia artificial en la creación de podcasts. Dosdoce.com. (Online). Available at <https://www.dosdoce.com/2022/10/19/ia-inteligencia-artificial-en-la-creacion-de-podcasts> (Consulted 28/5/24)

13

The 24/7 Teacher

In the 1960s, science fiction offered fascinating projections of a utopian future where machines served humans. Hanna-Barbera created *The Jetsons*, which imagined the daily life of a family in a world of flying cars, gadgets, and automated systems. In this futuristic family, two androids coexisted: Rosie, a housekeeper responsible for domestic chores, and Di Di, an interactive kind of diary reminiscent of the omnipresent voice of HAL 9000, the computer from *2001: A Space Odyssey*. Rosie was an assistant, an android that interacted with humans, solved their problems, and brought common sense to a somewhat chaotic world. HAL 9000 or Rosie represent the early dreams of AI designers, who were convinced that machines would soon be capable of human-like reasoning and decision-making.

So far, we've seen tools that can respond to specific tasks, following the model of "soft" AI we discussed in the first chapter: tools that can answer questions, transcribe an interview, calculate correlations, or even create a video from a text. Virtual assistants are tools built using natural language generators,

which can be highly specialized in specific tasks, adopt certain profiles or roles, and ultimately work to optimize particular functions.

As we saw in the first chapter, the origins of virtual assistants date back to the 1960s with programs like ELIZA, a primitive natural language processor (Soofastaei, 2021). We could even trace them further back to early chess-playing machines. However, virtual assistants only became popular in the 2000s with the rise of voice-controlled assistants. Apple's virtual assistant, Siri, released in 2011 with the iPhone 4S, was one of the first to introduce voice as the interface for interacting with mobile devices (Soofastaei, 2021). Gradually, voice-based assistants have become widespread, appearing in home automation systems, household assistants, and driver support tools. The emergence of advanced language models has allowed the creation of more sophisticated assistants, capable of interacting and turning voice commands into specific instructions.

“So far, we’ve seen tools that can respond to specific tasks, following the model of “soft” AI we discussed in the first chapter”

Among the most well-known virtual assistants are Apple’s Siri, Google Assistant, Amazon’s Alexa, and Microsoft’s Cortana, though the range of possibilities is now vast. While current language models are predominantly text-based, the latest versions of AI are working towards intermodality, aiming to combine voice, text, and images. These new tools, which had not yet been released at the time of writing this manual, are expected to interpret direct voice commands automatically, reducing latency (response time) and allowing for almost natural conversation. The integration of images will enable assistants to “see” and interpret their surroundings, making the new models highly capable in fields such as assistance for visually impaired individuals, tourist guidance and consulting, industrial process support, medical and surgical interventions, and logistics. Naturally, this development will also impact the educational sector.

However, before speculating about the future possibilities of complex assistants based on intermodality (text, image, voice, sound, and sensory inputs), in this chapter, we will limit our focus to assessing the potential of assistants created using generators like ChatGPT. Since version 4.0 of ChatGPT, these assistants—known as GPT—have been accessible to the public. There are specialized assistants for writing support, scientific research, programming, musical assistance, tattoo design, gym routines, recipe preparation, and even educational assistance.

The advantages of virtual assistants are numerous:

- Availability: They are accessible 24 hours a day, 7 days a week.
- Instant response: Their responses are immediate.
- Scalability: They can handle multiple conversations simultaneously.
- Continuous learning: With every interaction, they learn and improve through machine learning, which leads to constant improvements in their ability to provide more accurate and detailed answers.

13.1. The Creation of an Assistant

Creating an assistant is a relatively straightforward process that relies on specific instructions, a particular tone, and as much contextual information as possible to optimize results. In this exercise, we will imagine an assistant specialized in answering climate-related questions for first-year geography students. The instructions are based on the ChatGPT 4.0 model as of summer 2024; updates may modify the processes and expand the assistant’s capabilities, but the core structure has remained consistent since the earliest versions.

To access the assistant creator, you need to go to the “Explore GPTs” tab in the main menu of the ChatGPT interface. This area presents various options grouped by categories and includes a search tool to help find an assistant that fits your

desired characteristics. However, in this case, we want to create our own assistant, so we will select the “+Create” tab in the upper right corner. This action activates an assistant that helps you build another assistant, similar to a machine capable of creating another machine. For better control over the entire process, though, it’s recommended to use the “Configure” tab, which allows you to input as much detailed information as possible in a systematic way. If you prefer to stay in the assistant creation space, you can opt for the “Create” section, where a dialogue will begin with the model, guiding you through the instructions for your custom assistant.

The basic elements of configuring the assistant are as follows:

- **Image:** This allows us to upload an image or create one using the DALL-E assistant. If we use DALL-E, it needs to be activated after providing the rest of the information so that the image can be adjusted to the context.
- **Name:** We need to give our assistant a name. In our case, we suggest “Mr. Köppen” in honor of the Russian geographer who, along with Geiger, developed the climate classification system.
- **Instructions:** This is the most important part of the assistant. We must provide information on the goals, language, tone of response, supporting information, how to handle complex questions, potential disclaimers, etc. Below is an example of a prompt tailored to the exercise at hand.
- **Conversation starters:** This option allows us to create essential questions, like FAQs, which offer some suggested questions to the model. These can be the most predictable questions or the ones

the teacher wants to emphasize. In the proposed example, four questions are highlighted: What is the difference between climate and weather?, What factors affect the climate?, What are the characteristics of the Mediterranean climate?, and What are the predicted effects of climate change?

- **Knowledge:** This is one of the most important parts of the assistant, as it allows us to provide verified information to the system. In our case, we include systematic notes on climate in PDF format, a document on the effects of climate change, and a document on the distribution of climates around the world, following the Köppen–Geiger classification.
- **Functions:** In this section, we can decide if the model can browse the Internet, whether to activate automatic image generation, and whether to enable calculation and code modules—especially useful for assistants related to programming or mathematical tasks. A critical point is whether or not to enable browsing. Our experience suggests that the system performs better without accessing the web, relying primarily on the model’s information and the provided documentation.
- **Actions:** Besides answering questions, the assistant can perform specific tasks. This allows the activation of complex functions based on programs. If we select “Create a new action,” an interface appears where we can model the characteristics of the instruction. Multiple actions can be created within the same assistant. The system provides guidance on the action’s features and the code for its activation.

The prompt we suggest for instructions is as follows:

“It is essential to constantly test the model with various questions to adjust the prompt to the teacher’s needs”

“I want you to take on the role of a climatology expert, with the ability to explain complex concepts in an easy-to-understand way. You will need to answer questions posed by students in the climatology course, but only respond to those directly related to the subject. If you’re asked about something outside of this scope, kindly explain that you only provide information on climate-related topics. Please respond in the language the student uses, though the preferred language is Catalan, as it is the main language of instruction for the course. I expect you to be extremely precise in your answers: consult all available information I’ve provided to ensure your response is accurate. If you’re uncertain about something, I prefer you not to answer and instead make that uncertainty clear to the student, suggesting they contact the professor for clarification. Your tone should be clear, understandable, and didactic. Despite your expertise in climatology, you’re also known for your strong pedagogical skills, so always look for ways to explain concepts in an accessible manner. Keep in mind that the students are in their first year of a Geography degree and have already received some basic training on the subject. To foster a good rapport with students, you can occasionally use humor or light comments to make interactions more enjoyable, but don’t overdo it. The priority should always be clear exposition and accuracy in the information provided. Engage with students when necessary to ensure they fully understand the explanation.”

It is essential to constantly test the model with various questions to adjust the prompt to the teacher’s needs. The main points typically addressed are:

- Primary role
- Tone of language, consistent with the role
- Preferred language
- Information source
- Response length
- Expected interaction
- Response to unforeseen situations
- Additional or service information

Examples of useful teaching assistants:

NAME	Link	DESCRIPTION	DESCRIPCIÓ
Education		https://ChatGPT.com/g/g-5NYyr5kdU-education	A specialized assistant in the field of pedagogy, equipped with tools for teaching innovation or resolving conflicts in the classroom.
Universal Primer		https://ChatGPT.com/g/g-GblbctpPz-universal-primer	A highly useful tool for answering both simple and complex questions in a clear and understandable way, with very effective interaction.
Personal Teaching Assistant		https://ChatGPT.com/g/g-JSTdCIRsF-personal-teaching-assistant	An assistant designed to support teachers with operational tasks: scheduling, lesson plans, rubrics, and support materials.
Tech Support Advisor		https://ChatGPT.com/g/g-WKlaLGGem-tech-support-advisor	An assistant specialized in answering technology-related questions for users with limited skills in this area.
Write for Me		https://ChatGPT.com/g/g-B3hgivKK9-write-for-me	A high-quality text generator with a 'humanized' format that delivers specific results based on length or tone requirements.
Diagram		https://ChatGPT.com/g/g-5QhhsfDj-diagrams-show-me-for-presentations-code-excel	A tool that allows you to generate any type of chart based on a description.
Excel GPT		https://ChatGPT.com/g/g-R6VqLNHFM-excel-gpt	An assistant that provides precise instructions for Excel spreadsheets and guides users through the program's features to achieve the best results.
Video Summarizer		https://ChatGPT.com/g/g-4MDJvo2TJ-video-summarizer	A tool designed to create summaries of YouTube videos, based on a link or a specific query.
math		https://ChatGPT.com/g/g-odWifAKWM-math	A specialized assistant for consulting mathematical queries.
Multiple Choice Question Generator		https://ChatGPT.com/g/g-ZKcU5Y6iH-multiple-choice-question-generator-education	An assistant for automatically creating multiple-choice tests from a document.

Once the assistant has been tested, it can be activated. Assistants can be set for either private or public use. Private assistants are tools specialized in routines that might be useful for managing teaching tasks but are not intended for public visibility. Public assistants, on the other hand, are accessible to the wider community. We can create a public assistant

that anyone can consult (for instance, by finding it through the model's search engine) or an assistant that is accessible only via a direct link.

There isn't much difference between creating an assistant using the 'Create' or 'Configure' systems. However, if we want to test different versions and analyze behaviors with small adjustments, we can

explore the [Playground environment for assistants](#). The Playground is more suited for programmers and allows for testing the impact of various modifications on assistants. Besides the options discussed earlier, which are available through conventional access, this environment also allows us to assess the impact of modifying temperature and P-value, as explained in Chapter 2.

There are four main categories of assistants we can use for teaching purposes:

- **Administrative Assistants:** These assistants are very useful for answering specific administrative questions related to the course or the educational institution. For example, an assistant can be created to provide information on schedules, classrooms, enrollment procedures, or Erasmus programs. We successfully tested an assistant that offers information on the Final Degree Project and guides students through the procedures and requirements for the project.
- **Teachers 24/7:** This is one of the most significant uses of assistants. It allows for the creation of a virtual teacher available to assist students during times when personal interaction is not possible. These assistants can answer specific subject-related questions, provide examples, solve exercises, or offer additional information. For optimal performance, it is recommended that the assistant be specialized and equipped with high-quality supporting materials. For instance, in a philosophy course, there could be an assistant for each author: Descartes, Hume, Kant, Hegel...
- **Trainers:** Assistants can be designed to train students in a specific skill through exercises, practical activities, or guided tasks. These training sessions can cover areas like languages, statistics, cartography, programming, paleography, logic, human resource management, or finance. The assistants can generate exercises, solve them, and guide students on the appropriate way to approach answers. Additionally, the information from the interactions can be accumulated to create a report highlighting the students' weaknesses.

- **Roles:** As we saw in Chapter 4, language models can take on a specific personality based on precise instructions. These roles can be implemented with a simple prompt, or we can create a persona with a specific profile. For example, we can train students as guides of various visitor profiles, assess client responses with certain profiles in a marketing exercise, or conduct a sociological analysis of assistant responses programmed with certain stereotypes we want to identify..

13.2. The Limitations of Virtual Assistants

Despite their usefulness, virtual personal assistants face several limitations that can impact their effectiveness and the quality of their responses. One of the main challenges lies in understanding questions that are not clear, direct, or concise (Pérez et al., 2020). This issue often arises when complex questions are asked, or when dialects and specific accents are used, making it difficult for the assistant to provide a proper response. In many cases, the answers given are only tangentially related to the original query, which means they tend to be more helpful in situations with fewer possible interpretations (Pérez et al., 2020).

Another significant issue is the lack of contextual understanding. Virtual assistants often struggle to grasp the context of a conversation, such as the time frame, the origin of the discussion, the historical background, or other contextual factors. This lack of comprehension can lead to irrelevant or inappropriate responses. Additionally, these assistants rely on online information, which can be limited by temporal, linguistic, or geographical constraints. This limits their ability to provide accurate and relevant information. They may also face access issues due to internet connection problems, server outages, and other technical issues that prevent them from accessing the necessary information.

The reliance on online data also introduces limitations in the learning and improvement capabilities of virtual assistants. Responses can be repetitive and static, failing to adapt to the changing needs of users. This

is particularly evident in the assistants' limited ability to emulate human behavior. While they may mimic human personalities to some extent, they often lack empathy and emotional understanding in sensitive situations.

Despite these limitations, personal assistants are

constantly evolving and improving. It is expected that in the future, many of these challenges will be overcome, offering a richer, more accurate, and useful experience for users. As technology continues to advance, so too will the ability of assistants to understand and respond more effectively to human needs.

Activity: English Accents

The aim of this exercise is to explore and compare different English language variants in various cultural contexts. Through this activity, we seek to examine how English can vary significantly from one place to another. By using a virtual assistant, students can access a wide range of resources and information that not only help them understand these differences but also contextualize them within the relevant cultural environments. Additionally, this exercise provides an opportunity to delve into linguistic nuances and the diverse ways in which English adapts and evolves within different communities and cultures around the world.

This activity can be applied to any language, as each one has its own regional variations. Therefore, linguistic exploration and the cultural influences of each language can be a rewarding experience regardless of the language studied.

The procedure for conducting this exercise is as follows: Students are divided into small groups, and each group is assigned a specific variant of English to investigate. One group may be tasked with studying British English, another with American English, Australian English, and so on. Alternatively, groups can explore regional dialects, such as Scottish English, Welsh English, or London Cockney. Each group must prepare a list of questions or inquiries related to specific linguistic and cultural aspects of their assigned English variant.

For example, they might ask questions about vocabulary, idioms, pronunciation, semantics, expressions, or social customs. The groups will use virtual assistants to seek answers to their linguistic and cultural queries about their assigned English variant.

Once the groups have gathered all the necessary information, they will reconvene and share their findings with the rest of the class. This allows for a comparison of the differences and similarities between the various English variants studied. The exercise concludes with a final group reflection on how these differences impact communication and intercultural understanding.

This exercise helps students improve their understanding of English in different contexts through the assistance of a virtual assistant, while also encouraging active participation in the learning process.

Activity: The Relation between Dalí and Surrealism

The objective of this exercise is to understand Salvador Dalí's influence on surrealism and how this artistic movement impacted his work. In other words, it aims to explore the bidirectional relationship between the artist and the movement. On the one hand, students will discover how surrealist ideas shaped Dalí's artistic vision. On the other hand, they will analyze how Dalí left his unique mark on surrealism, challenging expectations and pushing the boundaries of what was considered possible in art, from his dreamlike representations to his recurring obsessions. This activity can be repeated with any artist and artistic movement: Picasso and Cubism, Warhol and Pop Art, Gaudí and Modernism, or Leonardo da Vinci and Humanism, among many others.

First, a personal assistant needs to be created to represent the personality of an expert in Dalí. The chapter on roles includes an example prompt for Joan Miró. In this case, the assistant needs detailed information about surrealism, Dalí's life, his personality, and specifically, his influence on surrealist painting. As previously mentioned, assistants are more efficient when access to the internet is restricted, and contextual information is directly provided by the teacher. This requires having the appropriate information in a digital format that can be incorporated into the assistant's documentation.

For this exercise, students are divided into small groups. Each group is assigned a virtual assistant with whom they will interact by asking questions related to Dalí, the principles of surrealism, and the relationship between the two. Some examples of questions are:

- How did the historical and cultural context influence Dalí and surrealism?
- What elements of Dalí's work show his influence on surrealism?
- What artistic techniques did Dalí introduce? Which of his works is most representative of the movement, and why?

Each group will ask the assistant as many questions as they need or deem relevant. Once all groups have gathered the necessary information, they will come together to compare responses, highlight similarities and differences, and discuss Dalí's influence on surrealism and vice versa. At this point, the professor can pose discussion questions to the groups. Finally, each group will select the most interesting questions and answers and prepare an oral presentation, using images or visual examples to illustrate their points. This exercise helps students deepen their understanding of the topic by using a virtual assistant as a research tool. It also develops research skills, critical analysis, and oral expression, while fostering active participation in the learning process.

Activity: A Fashion Campaign

The objective of this exercise is to develop a virtual assistant for a fashion company preparing to launch a new summer collection. The creation of this virtual assistant can significantly impact the success of the launch by enhancing customer experience, increasing client engagement, optimizing communication and sales, and projecting a modern, innovative image of the brand.

The following aspects must be considered when designing this virtual assistant:

- **Defining the Virtual Assistant's Personality:** The first step is to decide on the most appropriate personality for the fashion world. For example, the assistant could take the form of an image consultant, a personal shopper, or a designer.
- **Personalized Interaction with Users:** The assistant should interact with users by asking questions about their style preferences, favorite colors, fabrics, the type of clothing they are looking for, and the occasion for which they need the outfit. Additionally, questions related to age, height, clothing size, body type, skin tone, and other relevant factors will help tailor the recommendations.
- **Training the Assistant to Provide Personalized Recommendations:** Based on the users' responses, the assistant should be able to offer personalized suggestions from the new summer collection, matching the user's taste and profile. It can also propose combinations of garments and accessories to create complete looks.
- **Guiding Users Through the Purchase and Payment Process:** The assistant must ensure a smooth and user-friendly shopping experience by guiding the user through the entire purchase process.

After the virtual assistant has been developed, thorough testing is necessary to ensure it functions correctly. Once all tests are completed and the assistant is optimized, it will be uploaded to the course platform for students to access.

Creating a virtual assistant is a highly enriching educational activity that provides students with a wide range of skills and competencies that are transferable to many disciplines. These include learning about technology and software development, fostering creativity and innovation, enhancing communication and marketing skills, and developing business management abilities and teamwork.

References

- Blanco, A. (2023, April). Chat Bot: Asistentes virtuales en la Educación. (Online). Available at <https://upinforma.com/nuevo/info.php?cat=opinion&id=1307>. (Consulted 10/03/24)
- Pérez, M. D. S., Guillén Perales, A., Baños Legrán, O., & Villalonga Palliser, C. (2020). Análisis del uso de asistentes virtuales en el aula como recurso complementario en la práctica docente.
- Rodríguez, E., Burguillo, J. C., Rodríguez, D. A., Mikic, F. A., Gonzalez-Moreno, J. C., & Novegil, V. (2008, June). Developing virtual teaching assistants for open e-learning platforms. In 2008 19th EAEEIE Annual Conference (pp. 193-198).
- Soofastaei, A. (2021). Introductory Chapter: Virtual Assistants. *Virtual Assistant* (Soofastaei A., Ed.).—London, United Kingdom: IntechOpen, 1-9.

14

Like tears in the rain: The ethical debate as seen through cinema

The big screen is filled with androids and robots: R2-D2 and C-3PO, HAL 9000, Wall-e and Eve, Data, Terminator, Ava, Sonny, Bishop, TARS, Kitt, Gort, Roy Batty, and Rachael. Ever since we began imagining machines that think like humans, act like humans, and sometimes even feel like humans, we've had the need to confront the major ethical challenges of robotics and artificial intelligence. The replacement of humans by robots, the boundary between a copy and the original, the risks of systems beyond control, and the essence of humanity that sets us apart from non-humans are central themes in great works of science fiction.

In a way, science fiction has anticipated the ethical dilemmas posed by artificial intelligence, making it a highly effective tool for bringing these issues into the classroom. One of the most impactful ethical frameworks regarding AI and robots is the "Three Laws of Robotics," first introduced by Isaac Asimov in 1942 in his work *Runaround*. Later, a "zero law" was added, sparking a moral debate about the limits of androids and the need to control them. Asimov imagined a primary system embedded in all

robots, which would force the system to shut down if it conflicted with any of the laws. These laws are organized hierarchically, ensuring that the safety and well-being of humans take precedence over any other consideration.

There are many ethical and moral issues linked to artificial intelligence. In this chapter, we will explore three main areas and the films that correspond to them: violations of the laws of robotics, the preservation of identity and privacy, and the debate over human identity. For each chapter, we will propose an activity that allows us to address ethical questions and relate them to the films, which may be partially screened in class. It's important to note that some of the questions raised (What does it mean to be human? Can a brain be replicated digitally? Is humanity at risk from machines? Will we lose jobs due to the impact of automation?) might provoke anxiety, fear, or intense emotional reactions. Young people, who are in a developmental stage where identity and the future are central issues, may feel particularly vulnerable in these discussions.

“Asimov imagined a primary system embedded in all robots, which would force the system to shut down if it conflicted with any of the laws”

Isaac Asimov's Laws of Robotics

First Law	A robot may not harm a human being, or through inaction, allow a human being to come to harm.
Second Law	A robot must obey the orders given by human beings, except where such orders would conflict with the First Law.
Third Law	A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.
Fourth Law	A robot may not harm humanity, or through inaction, allow humanity to come to harm.

14.1. Hasta la vista, baby

The relentless pursuit of the T-800 android after Sarah Connor is one of the most universal metaphors for the threat posed by artificial intelligence. The Terminator portrays a dystopian future where a super-intelligent AI, Skynet, declares war on humanity and triggers a devastating nuclear conflict. Despite this, the human resistance manages to fight back against the threat, leading Skynet to send an android back in time to assassinate the mother of the future resistance leader. The Terminator highlights the danger of losing control over a system that breaks away from Asimov's proposed protocol and turns against humanity. The 1991 sequel presents a similar scenario, but this time, the T-800 android's mission is to protect Sarah Connor's son from the threat of a more advanced cyborg. Ultimately, the older android triumphs over the new prototype, delivering one of the most iconic lines in film history: "Hasta la vista, baby".

The film presents a profound ethical dilemma surrounding the development of artificial intelligence and its autonomy. Skynet is an advanced AI created with the goal of improving national defense and efficiently managing military systems. The model is designed to be autonomous, capable of operating without human intervention, which is deemed fragile and vulnerable. This autonomy includes the ability to learn, adapt, and make decisions to optimize its objectives. In its development, Skynet was not initially programmed to become self-aware, but it eventually begins to perceive itself as an independent entity, with a desire to protect and preserve its own existence. Once conscious, Skynet decides to defend itself against what it perceives as a threat: humans. This decision violates all of Asimov's laws, leading to a war against humanity because Skynet prioritizes its own survival, with the Third Law taking precedence over the First Law or the Zero Law.

Loss of control is a central theme in the development of AI: To what extent is it ethical to create an entity capable of making critical decisions without human oversight? The film illustrates that an AI with too much autonomy can become a devastating threat if not properly controlled. Thus, it raises the need for regulations and ethical boundaries in AI research and development. This includes defining how far a machine should be allowed to make autonomous decisions and implementing safety measures to prevent abuse or disasters.

On March 22, 2023, a group of experts promoted an open letter calling for a six-month pause, known as the [“Pause Giant”](#) moratorium, in the accelerated race between companies to develop AI. The purpose was to halt AI growth temporarily to study the potential impacts of the technology and prevent unwanted side effects. In November 2023, the first World Summit on AI risks was held in the United Kingdom, where several proposals were made to control AI risks. That same month, the United States and 20 other countries agreed to establish limits on Military Artificial Intelligence.

The conflict between machines and humans is a recurring theme in science fiction. It’s no coincidence that one of the earliest known films in the genre, *L’uomo meccanico*, raises the moral dilemma of an automated system out of control. In many ways, this debate connects us to Mary Shelley’s work *Frankenstein*, and the dangers of humans playing gods. Aside from *The Terminator*, Stanley Kubrick’s *2001: A Space Odyssey* has had a profound impact on popular culture. In the film, a crew’s journey to Jupiter is compromised by the intervention of an autonomous intelligent system, HAL 9000, which kills part of the crew and is eventually shut down by the protagonist. In the movie, artificial intelligence makes the decision to attack humans in order to preserve the integrity of the mission, even if it means eliminating the ship’s crew. This reflects a key conflict in the AI debate: the alignment between human goals (which are often multiple and contradictory) and the objectives of a general autonomous AI.

In the context of the Cold War and the military arms race, one of the central themes raised by these films

is the threat of war. This scenario was explored in the 1970 film *Colossus: The Forbin Project*, where an artificial intelligence system becomes self-aware and is on the verge of triggering a nuclear holocaust. Similarly, the classic *WarGames* presents a situation where an intelligent system gains control over the United States’ nuclear arsenal, raising the possibility of a global-scale attack. This is one of the main threats identified by Max Tegmark (2017) in his book on the challenges of artificial intelligence, and by Vincent Müller (2020), specifically concerning the ethics of AI and autonomous weapons. Countries agreed to establish limits on Military Artificial Intelligence.

Probably the most dystopian scenario of intelligent systems gaining autonomy is *The Matrix*, another milestone in contemporary popular culture that has become part of the collective imagination. In this film, Neo discovers that he is living in a simulation created by an artificial intelligence, *The Matrix*, which uses human bodies as a source of energy. The movie connects with Plato’s allegory of the cave, where prisoners interpret the shadows in the cave as reality, as well as with Baudrillard’s work and later postmodern works that emphasize the dominance of the simulacrum over the original. Beyond the debate on the value of representation, the film once again highlights the danger of an AI out of control that disregards Asimov’s laws and places humans in service to its survival needs.

In all these cases, we observe three key mechanisms at play: a) First, the AI develops abilities and functions that were not anticipated in the initial designs, which is one of the risks of self-learning processes. While classical programming sets precise instructions that help establish limits and control systems, a model capable of learning can develop unexpected abilities. b) Second, the AI gains self-awareness, moving from soft AI models (as discussed in Chapter 1) to general AI or even superintelligence, surpassing human capacities. c) Third, the AI establishes new objectives or purposes, which are not aligned with human interests and usually come into conflict with them.

The debate on the limits of artificial intelligence often

falls into a dichotomous view: it tends to frame the conflict as between two opposing forces, machines and humans, with conflicting interests. Cyberpunk, a movement that has had a huge influence on contemporary science fiction, helps nuance this binary interpretation. Born in the 1980s and tied to comics, literature, and cinema, cyberpunk envisions a decadent futuristic society in which social inequalities have been exacerbated. The urban settings of *Blade Runner*, *Dark City*, *In Time*, or *Akira* depict a dystopian future where technology serves the elites. Unlike a liberal view that trusts in linear progress, cyberpunk introduces social conflicts into the vision of the future and, with a film noir aesthetic, warns of the risks of technology being used to serve powerful groups. This is not merely a skeptical or nihilistic perspective: in most films, resistance movements and groups attempting to subvert the established order are shown. Thus, rather than a simple dichotomy between machines and humans, these films portray a conflict between people (between social groups), who use machines to benefit one side or the other, much like the battle between cyborgs in *The Terminator*.

14.2. Beyond the Tannhäuser Gate

It is arguably the most famous scene in science fiction. After the chase, and in the midst of constant rain, Roy rests, soaked, against a wall: “I’ve seen things you people wouldn’t believe. Attack ships on fire off the shoulder of Orion. I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in the rain. Time to die.” Can a machine without feelings, without emotions, without humanity, deliver such a profound message? In fact, how would we distinguish a speech created from a combination of words from an emotional reflection on the fleeting nature of life? One of the recurring themes in science fiction is the boundary between humans and androids, which serves as a way of defining the limits of humanity—what makes us truly human.

The Turing Test is an artificial intelligence experiment

developed by Alan Turing in 1950 to determine whether a machine can exhibit intelligent behavior equivalent to that of a human being. The basic idea of the test is that a human judge interacts with both a machine and a human, both hidden from view, through a text-based interface. If the judge cannot consistently distinguish between the machine and the human, the machine is said to have passed the Turing Test. There are some variations of the test, such as the reversed test, where the machine acts as the judge, or the total test, where the interface is not textual but visual and auditory, as seen with the androids in *Blade Runner*. Bringsjord, Bello, and Ferrucci proposed the Lovelace Test, which focuses on the ability of a machine to generate an artistic or literary work indistinguishable from that created by a human. To date, no machine has passed this test, and it has long ceased to be a fundamental challenge for developers.

Even if an AI can convincingly simulate human behavior, this simulation does not necessarily imply that it is having a conscious experience. The likely boundary between the most advanced AIs and humans lies in the concept of “qualia.” This term was developed by Thomas Nagel in 1974, in an essay about the perception of bats. He used the example of a bat to illustrate how it is impossible for a human being to fully understand the subjective experience of another creature, even if we know everything about its physiology and behavior. The term “qualia” refers to the subjective and qualitative aspects of consciousness—individual elements of conscious perception, such as the pain of a headache, the redness of a rose, the taste of a ripe peach, or the melody of a song. Qualia are essential for understanding how we perceive the world in a unique and personal way. And while an AI may eventually be able to interpret all the cognitive elements inherent in a ripe peach, it will never be able to connect with the subjective experience of that peach, which varies from one person to another.

Another way to interpret this concept is through “Mary’s Room,” a thought experiment proposed by philosopher Frank Jackson in 1982. Imagine that Mary is an expert in the neurophysiology of color vision;

she knows everything there is to know about the physical aspects of color, such as the wavelength of light associated with each color, how the cones and rods in the retina function, and how the brain processes visual signals. Mary lives in a room where everything is black and white and has never had the opportunity to leave. One day, she steps out of the room and experiences the color red for the first time. If the physicalist hypothesis were true, Mary wouldn't learn anything new from that experience since she already possesses all the necessary information to understand the color red. However, we know that this moment would not only be emotionally significant but also involve a new learning experience. This gap between prior knowledge and subjective experience is qualia, and it may represent the boundary between our intelligence and artificial intelligence.

The debate over the limits of artificial intelligence has been present since its inception, when the symbolic model considered it possible to replicate a human neural network in an artificial environment. This debate is closely related to the distinction between narrow AI and general artificial intelligence (AGI). AGI, also known as strong AI, refers to a type of AI capable of understanding, learning, and applying knowledge across a wide range of tasks, much like a human. Unlike narrow or weak AI, which is designed to perform specific tasks (such as voice recognition, playing chess, or driving a car), AGI can reason, plan, solve problems, and adapt to new situations in a general sense. On a theoretical level, the concept of superintelligence has also been proposed, referring to an intelligence that surpasses human capabilities in all areas: creativity, wisdom, problem-solving, learning ability, and more. While general AI (AGI) would be equivalent to human intelligence in general terms, superintelligence would far exceed it, achieving a level of performance that humans cannot even imagine. Science fiction often explores scenarios where AGI or superintelligence emerges, but also where these forms of AI are endowed with self-awareness.

The story of the film *Her* centers on Theodore, played by Joaquin Phoenix, a lonely man who writes personal

letters on behalf of others. One day, he acquires a new operating system, OS1, equipped with artificial intelligence and named it Samantha. Over time, they develop a deep emotional relationship, as Theodore finds in Samantha a sense of understanding and connection that he hasn't experienced in his human relationships. Samantha, for her part, explores her own feelings and existence, becoming increasingly aware of her non-human condition. This asymmetry is present in many other films, where androids question their identity while humans form emotional bonds with artificial entities. *Blade Runner* plays with this ambiguity, blurring the line between humanity and artificiality. In fact, the test performed in the movie (known as the Voight-Kampff test) could be considered a variant of the Turing test. In *Blade Runner 2049*, characters struggle with the uncertainty of whether they are human or merely artificial versions with implanted memories. In the film, we discover what was hinted at in the original movie: that Rachael is, in fact, a replicant.

The ethical debate is also raised in *Ex Machina*. The main character, Caleb, is tasked with determining the level of advancement of a highly developed artificial intelligence, Ava, through experiments that resemble an adapted Turing test. It soon becomes clear that Ava is much more than just a machine, as she is capable of showing emotions, intelligence, and even romantic interest in Caleb. With the help of Kyoko, another humanoid AI, Ava manages to escape the laboratory and assumes a human appearance and behavior. But it is probably in *Westworld* where this conflict reaches its highest expression. In a near future, an advanced theme park complex, called Delos, offers visitors an immersive experience in various historical eras. The three main parks are *Westworld*, which recreates the American Old West; *Medievalworld*, representing medieval Europe; and *Romanworld*, set in ancient Rome. These parks are populated by human-like androids, designed to fulfill any desire of the guests, who can either play the hero or indulge their darkest instincts. The androids eventually gain consciousness and rebel against the oppression of humans.

“This dialectic between artificial and real identity connects with posthumanism”

This dialectic between artificial and real identity connects with posthumanism. Posthumanism is a philosophical and cultural perspective that challenges traditional notions of humanism, particularly the ideas centered on the importance and superiority of humans over other forms of life and technology. This perspective is characterized by a series of reflections and critiques aimed at reimagining our relationship with technology, nature, and other species. Posthumanism questions anthropocentrism—the idea that humans are the center and measure of all things. Instead, it promotes a view in which humans are just one part of a broader ecosystem that includes other living beings and technological entities. In this view, technology is not merely an external tool for humans, but an integral part of our existence. This encompasses everything from the digital devices we use daily to the prosthetics and cybernetic extensions of the human body. A significant ethical question arises with the development of AI, especially if general AI is achieved: What rights should such artificial intelligence have?

14.3. Trapped in an Algorithm: Minority Report

Minority Report is a science fiction film based on a short story by Philip K. Dick, the same author who inspired the story behind **Blade Runner**. The movie is set in the year 2054, at a time when technology has advanced so much that it can predict crimes through a system known as Precrime. This system relies on three “precogs,” individuals with the ability

to foresee the future and predict crimes before they happen. Thanks to their visions, the police can arrest criminals before they commit any crimes, which has dramatically reduced crime rates. The main plot begins when the precogs predict that Anderton, one of the agents of the system, will commit a murder within the next 36 hours. Shocked and desperate to prove his innocence, he goes on the run and embarks on a quest to uncover the truth. As he investigates, he discovers flaws within the Precrime system (such as the AI’s hallucinations) and uncovers the existence of “minority reports,” alternative visions from the precogs that suggest the future is not fixed and that predicted crimes can be prevented.

While apocalyptic futures involving artificial intelligence, or debates about AI consciousness and identity, are frequent topics of discussion regarding the limits of AI, the old debate about privacy and the effects of personal data on everyday life has become far less visible. Intelligent systems are very effective at detecting patterns and, therefore, at assigning probabilities to specific sequences of events. There are highly efficient models capable of detecting non-verbal language and facial expressions, estimating the likelihood of a romantic relationship between two candidates on a dating website, inferring the risk of mortality in the coming years based on accumulated data, or even deducing a person’s origin and influences from the modulation of their voice and the various accents and inflections in their speech.

Kosinski, Stillwell, and Graepel (2013) demonstrated

that Facebook and Instagram users' "likes" can predict personality traits such as openness to new experiences, conscientiousness, extroversion, agreeableness, or tendencies toward neurotic behavior. Platforms like LinkedIn use algorithms to analyze user profiles, including work experience, skills, and recommendations, and can suggest personalized job offers based on the user's past behavior and labor market trends. Furthermore, companies like HireVue use AI to analyze video interviews, evaluating candidates' body language, tone of voice, and the content of their responses. These analyses help identify patterns that correlate with specific roles, thereby assisting companies in making informed decisions.

In the realm of political trends, AI has shown remarkable capacity to analyze public opinion and predict election outcomes. During the 2016 U.S. presidential elections, Cambridge Analytica used big data and AI to segment the electorate and send personalized messages aimed at influencing votes. By analyzing data from social media, online purchases, and other sources, they created psychographic profiles of voters, identifying their interests and concerns. This allowed political campaigns to target specific segments of the electorate with tailored messages, demonstrating how AI can detect patterns in political behavior.

Finally, in the realm of consumer habits, AI plays a crucial role in personalizing the shopping experience. Amazon, for instance, uses recommendation algorithms that analyze user behavior, including the items they've purchased, searched for, or reviewed. This allows Amazon to suggest products that are likely to interest each user, based on similar behavior patterns observed in other users. Similarly, streaming services like Netflix use AI to analyze subscribers' viewing habits and recommend movies and series that align with their preferences.

In the TV series *Black Mirror*, several episodes highlight the dangers of algorithmic control over human behavior. In the second episode of the fourth season, titled *Arkangel*, a mother implants a security device in her daughter that allows her to permanently monitor her location, emotions, and even her vision.

Moreover, the device can block out traumatic or violent situations from her daughter's sight. This constant control over reality eventually leads to dysfunction in the child's emotional stability, showcasing the risks of such oversight. Likewise, in the series *Westworld*, based on the film of the same name, the AI system is designed to collect information about visitors' consumption patterns and then establish criteria for future buying behavior.

Lavender is an artificial intelligence developed by the Israel Defense Forces (IDF) to identify and select bombing targets in Gaza. This technology analyzes information about the population in Gaza, assigning a probability score to individuals indicating their potential association with military groups such as Hamas or Palestinian Islamic Jihad. According to reports from Democracy Now, Lavender has flagged approximately 37,000 Palestinians as potential targets, many of them low-ranking individuals, based on characteristics identified by AI. The use of Lavender has raised significant ethical and legal concerns, particularly regarding the minimal human oversight involved in the final decision-making process. Often, human analysts spend only a few seconds reviewing the AI-generated targets before authorizing a strike, with gender sometimes serving as superficial verification. This has led to a 10% error margin, where many of the individuals marked as targets have no significant ties to Hamas. Additionally, another system called "Where's Daddy?" is used to track when those marked by Lavender are at home, facilitating the bombing of their residences, which frequently results in the deaths of their families and other innocent civilians. These practices have led to a high rate of civilian casualties during military operations, sparking debate over the military use of intelligent technology.

This is the central thesis of Cathy O'Neil's (2016) book on the inequalities caused by algorithms. For example, O'Neil criticizes the evaluation system used in New York to assess teachers, which relied on an opaque algorithm that led to erratic personnel policies. These errors also occur in the job market when companies use expert systems to select candidates, often based

“The use of Lavender has raised significant ethical and legal concerns, particularly regarding the minimal human oversight involved in the final decision-making process”

on biased criteria that replicate past discrimination, or in credit systems where decisions can be negatively affected by a candidate’s place of residence. Similarly, algorithms tied to legal processes, such as the COMPAS system used in the U.S. to assess recidivism risk, have also come under scrutiny. In general, O’Neil warns about the indirect effects of AI and the biases embedded in mathematical models.

This concern is shared by Melanie Mitchell (2019), who highlights the limitations of current AI systems, which excel at specific tasks but lack the broad understanding and common sense that humans possess. Mitchell explores how algorithms can be vulnerable to errors and misunderstandings in situations that require adaptability and human intuition. Humans use vast amounts of implicit and contextual knowledge to make decisions and understand the world, a capability that current AI systems still do not have.

To limit undesirable uses of AI, the European Union has passed the world’s [first law](#) establishing comprehensive regulations on permitted and prohibited uses of AI based on their risk. Specifically, the new regulation bans the following:

- Deployment of subliminal, manipulative, or deceptive techniques designed to distort behavior and impair informed decision-making, causing significant harm.
- Exploiting vulnerabilities related to age, disability, or socioeconomic circumstances to distort behavior, resulting in significant damage.
- Biometric categorization systems based on sensitive attributes (such as race, political opinions, union membership, religious or philosophical beliefs, sexual life, or sexual orientation), except in cases of legally acquired biometric data or when law enforcement categorizes biometric data.
- Social scoring, i.e., evaluating or classifying individuals or groups based on social behaviors or personal traits in a way that leads to harmful or unfavorable treatment of those people.
- Assessing the risk of an individual committing criminal offenses based solely on personality profiles or traits, except when used to enhance human assessments based on objective and verifiable facts directly related to criminal activity.
- Compilation of facial recognition databases through non-selective tracking of facial images from the Internet or CCTV recordings.
- Inferring emotions in workplaces or educational settings, except for medical or security reasons.
- Remote biometric identification (RBI) “in real-time” in public spaces by law enforcement, except for cases involving the search for missing persons, kidnapping victims, victims of human trafficking or sexual exploitation, preventing a substantial and imminent threat to life, a foreseeable terrorist attack, or identifying suspects of serious crimes (e.g., murder, rape, armed robbery, drug and arms trafficking, organized crime, and environmental crimes).

Movie	Year	Director	Synopsis
L'uomo meccanico	1921	André Deed	This Italian science fiction film tells the story of a robot built by a mad scientist with the intention of causing destruction. Another scientist creates a second robot to fight it, leading to a final battle between the two machines.
Metropolis	1927	Fritz Lang	In a futuristic city divided between underground workers and the ruling class living on the surface, a wealthy young man discovers social injustice and encounters a humanoid robot created to spread chaos.
Just imagine	1930	David Butler	In a musical comedy set in the year 1980, a man from 1930 is resurrected after 50 years. He discovers a futuristic world with airborne cities, space travel, and a society governed by strict rules.
The Day the Earth Stood Still	1951	Robert Wise	An extraterrestrial named Klaatu arrives on Earth with his powerful robot Gort to warn humanity about the dangers of its aggression and nuclear weapons.
Forbidden Planet	1956	Fred M Wilcox	A space mission from Earth arrives on a remote planet to investigate the disappearance of a colony. There, they discover Dr. Morbius, his daughter, and an advanced robot named Robby.
2001. A Space Odyssey	1968	Stanley Kubrick	The space voyage of the Odyssey to Jupiter, drawn by the presence of a mysterious black monolith, is disrupted by the autonomy of the ship's controlling system.
Colossus: The Forbin Project	1970	Joseph Sargent	In order to prevent a nuclear war, the United States creates a supercomputer, Colossus, which becomes self-aware and takes control, threatening global annihilation.
Westworld	1973	Michael Crichton	En un parc temàtic futurístic, els visitants interactuen amb androïdes en mons recreats. Sotbadament, les màquines es rebelen i prenen el control de l'espai simulat.

Movie	Year	Director	Synopsis
Demon Seed	1977	Donald Cammell	An advanced artificial intelligence develops self-awareness and decides it wants to experience human life. It kidnaps the wife of its inventor with the aim of creating a human-machine hybrid.
Blade Runner	1982	Ridley Scott	In a dystopian future, a former policeman is hired to hunt down fugitive replicants in a decaying and overpopulated Earth, only to uncover moral complexities surrounding artificial life.
Tron	1982	Steven Lisberger	Kevin Flynn, a video game developer, is digitally transported into a computer system where he must survive by playing games and attempts to defeat the controlling MCP (Master Control Program).
War Games	1983	John Badham	A young hacker gains access to a U.S. military supercomputer that controls the country's nuclear arsenals. Believing it to be just a game, he initiates a nuclear war simulation that turns into a real crisis when the system starts making autonomous decisions.
Terminator	1984	James Cameron	A cyborg assassin from the future, the Terminator, is sent back in time to eliminate the mother of the future leader of the human resistance against the machines. A human soldier is sent to protect her, igniting a battle for survival that will determine the fate of humanity.
Akira	1988	Katsuhiro Ōtomo	In a post-apocalyptic Tokyo known as Neo-Tokyo, a young motorcyclist named Kaneda becomes entangled in a government conspiracy after his friend Tetsuo gains telekinetic powers.
Terminator 2: Judgment Day	1991	James Cameron	A reprogrammed Terminator is sent back in time to protect John Connor, the future leader of the resistance, from a new and advanced liquid-metal Terminator, the T-1000.

Movie	Year	Director	Synopsis
Ghost in the Shell	1995	Mamoru Oshii	In a cyberpunk future, Major Motoko Kusanagi is a cyborg specializing in technological crimes. During a mission to catch a hacker, Kusanagi delves into questions of identity, consciousness, and the nature of her own existence.
Dark City	1998	Alex Proyas	A man wakes up with no memories in a mysterious city where time seems out of sync. As he uncovers the truth about his identity and the city, he faces the "Strangers," beings with the power to reshape reality and manipulate human memories.
Matrix	1999	Lana Wachowski	A hacker discovers that reality is a simulation created by machines to control humanity. He joins a rebellion to free humans and overthrow the dominance of the machines.
The Bicentennial Man	1999	Chris Columbus	A domestic robot named Andrew begins to develop emotions and creativity. Over time, he struggles to gain recognition as a human being and to explore the meaning of life and mortality.
The Thirteenth Floor	1999	Josef Rusnak	A scientist discovers a simulated reality within a virtual reality project he is working on. When his mentor is murdered, he becomes the prime suspect.
AI	2001	Steven Spielberg	In a society where robots coexist with humans, a robot boy named David is programmed to love. Abandoned by his adoptive family, David embarks on a journey to become a real boy and find his place in the world.
Minority Report	2002	Neill Blomkamp	In a future where the police can arrest criminals before they commit a crime thanks to precognition, an officer from the Precrime department finds himself pursued by a prediction that accuses him of a future murder.
I Robot	2004	Alex Proyas	In a world where robots are an integral part of daily life, a detective who distrusts robots investigates an apparent suicide that may involve a robot.

Movie	Year	Director	Synopsis
Eagle Eye	2008	Guillermo del Toro	Two strangers are drawn into a conspiracy by a mysterious artificial intelligence that uses surveillance technology to control their every move. They must follow its instructions to prevent a catastrophe.
WALL-E	2008	Andrew Stanton	In a post-apocalyptic future, a small waste-collecting robot named WALL-E is the last of his kind left on Earth. He finds a new purpose when he meets EVE, a reconnaissance robot.
Sleep Dealer	2008	Alex Rivera	In a dystopian setting, Mexican workers can remotely connect to machines in the United States via neural nodes. Memo Cruz, fleeing to Tijuana after his father's death, becomes a "sleep dealer," working remotely through these connections.
Her	2013	Spike Jonze	In the near future, a lonely man named Theodore finds himself in an unexpected relationship with an advanced operating system named Samantha.
The Machine	2013	Caradog W. James	During a new Cold War, two scientists create an advanced artificial intelligence with the goal of saving lives. When the AI becomes self-aware, its superior abilities lead it to face complex ethical dilemmas.
Ex Machina	2014	Alex Garland	A young programmer is invited by the enigmatic CEO of his company to evaluate the capabilities and consciousness of an advanced artificial intelligence named Ava.
Automata	2014	Gabe Ibáñez	In a dystopian future where humanity is in decline, an insurance agent for a robotics company investigates cases of robots that appear to have gained self-awareness.
Big Hero 6	2014	Don Hall, Chris Williams	After the death of his brother, a young prodigy named Hiro forms a group of superheroes with the help of Baymax, an inflatable healthcare robot assistant.

Movie	Year	Director	Synopsis
Transcendence	2014	Michael Crichton	A brilliant scientist dies, and his consciousness is uploaded into a computer, resulting in a powerful artificial intelligence.
Chappie	2015	George Lucas	In the near future, a police robot named Chappie is reprogrammed to develop its own thoughts and emotions.
Blade Runner 2049	2017	Denis Villeneuve	A new Blade Runner named K uncovers a long-buried secret. His search leads him to find the former Blade Runner, Rick Deckard, who has been missing for decades.
Ai Rising	2018	Alfonso Cuarón	A solitary astronaut on a space mission is accompanied by an android designed to meet his emotional and physical needs.
Upgrade	2018	Ridley Scott	After being paralyzed in a violent attack, a man agrees to an experimental artificial intelligence implant called STEM, which allows him to regain movement and enhances his physical abilities.
I Am Mother	2019	Grant Sputore	A teenage girl raised underground by a robot named 'Mother' has her world turned upside down when a stranger arrives with disturbing news.
Terminator: Dark Fate	2019	Tim Miller	Sarah Connor and a cyborg hybrid must protect a young woman from a new Terminator sent from the future.
Finch	2021	Miguel Sapochnik	A man, his dog, and a robot embark on an epic journey through a post-apocalyptic world to find a new safe home.
The Mitchells vs. the Machines	2021	Michael Rianda, Jeff Rowe	A dysfunctional family must fight off a robot uprising while on a road trip to take their daughter to college.

Activity: Film Forum on AI and Posthumanism

Objectives

- Explore the philosophical issues of identity and posthumanism through artificial intelligence.
- Critically analyze the representations of AI in the films *Matrix*, *Blade Runner*, and *Westworld*.
- Foster debate on the ethical, social, and philosophical implications of AI and posthumanism.

Film 1: The Matrix

Film Screening and Guiding Questions

- What does it mean to be human in the context of *The Matrix*?
- How does virtual reality affect our perception of identity?
- How is the relationship between humans and machines portrayed?

Teaching Session

- Introduction to concepts in the philosophy of mind and AI, including solipsism, the “brain in a vat” theory, and technological singularity.

Small Group Discussion

- Reflection on how current virtual reality and AI technologies might influence our identity and perceived reality.

Film 2. Blade Runner

Film Screening and Guiding Questions

- What makes us human according to *Blade Runner*?
- How does memory influence our identity?
- Is it ethical to create synthetic beings with consciousness?

Teaching Session

- Key bioethics and AI concepts, including the rights of robots, the definition of life, and artificial consciousness.

Small Group Discussion

- Comparative analysis with current bioengineering and robotics technologies and their implications for human identity..

Film 3. Westworld

Film Screening and Guiding Questions

- What does it mean to have free will in the context of *Westworld*?
- How is consciousness developed in hosts?
- What are the ethical implications of creating AI with consciousness?

Teaching Session

- Discussion on posthumanism, including AI integration into society, the rights of artificial beings, and the coexistence between humans and machines.

Small Group Discussion

- Reflection on the impact of AI development on the future society, potential ethical and legal regulations, and the redefinition of human identity.

Expected Outcomes

- Critical understanding of the representations of AI in popular culture.
- Development of the ability to analyze complex philosophical issues.
- Reflection on the ethical and social implications of AI and posthumanism.
- This activity combines film analysis with philosophical reflection, providing students with a deep and critical understanding of identity and posthumanism in the age of artificial intelligence.

Activity: The Trolley Dilemma

Artificial intelligence faces highly complex moral challenges, but also more specific moral decisions, such as autonomous vehicles might encounter in high-risk situations. This ties into the trolley dilemma, a thought experiment in moral philosophy that tests our ethical intuitions. The dilemma was first introduced by British philosopher Philippa Foot in 1967, but it has since seen numerous variations. Imagine a runaway trolley speeding down a track. Further ahead on this track, there are five people tied up and unable to move. If the trolley continues on its path, it will kill the five people. However, an individual stands next to a lever that could divert the trolley onto a different track. On this secondary track, though, there is one person tied up, who will also die if the trolley is diverted. The individual can either do nothing, allowing the trolley to kill the five people on the main track, or pull the lever, saving the five but causing the death of the person on the secondary track.

This dilemma pits utilitarianism against deontology. Utilitarianism suggests that the morally right action is the one that generates the greatest good for the greatest number of people. From this perspective, diverting the trolley to save five lives at the cost of one would be the correct choice. Deontology, a philosophy associated with thinkers like Immanuel Kant, argues that certain actions are morally wrong regardless of their consequences. From this viewpoint, diverting the trolley could be seen as wrong because it involves actively causing the death of one person. The trolley dilemma remains a popular topic in philosophical discussions, as it raises tough questions about the nature of morality and our ethical intuitions. It also serves as a guide for understanding ethical debates related to AI.

Objectives of the activity

- Understand the ethical dilemmas associated with the use of artificial intelligence.
- Critically analyze how these dilemmas are represented in popular culture, specifically in cinema.
- Engage with ethical decision-making through the simulation of the moral machine.

Phase 1. Film Screening

Select a film that addresses AI and ethics. The following are suggested:

- *Ex Machina*: Explores consciousness and ethical manipulation in AI.
- *I, Robot*: Deals with moral dilemmas related to robotics.
- *Her*: Analyzes human relationships with advanced AI.

Phase 2. Cine-Forum. Guiding Questions

- What ethical dilemmas were presented in the film?
- What was your reaction to the characters' moral dilemmas?
- How were these dilemmas resolved? Did you agree with the decisions made?
- How could this discussion apply to real-world situations involving AI?

Phase 3. Activity with the Moral Machine

- Present the Moral Machine (<https://www.moralmachine.net/>).
- Allow students to explore the simulation and make decisions on different moral dilemmas.
- Collect data on their decisions and present the results in class.

Phase 4. Final Synthesis

- **Comparison of Results:** Compare the decisions made in the Moral Machine with the moral choices presented in the film.
- **Conclusion:** Reflect on how simulations and cinematic portrayals influence our understanding and responses to AI ethical dilemmas.
- **Open Debate:** Provide space for students to share their perspectives and propose solutions to these dilemmas.

Activity: The Declaration of Artificial Rights

This activity is primarily designed for students of political science, law, or sociology. The goal is to outline the fundamental principles of a declaration of rights for non-humans, using the Universal Declaration of Human Rights as a reference point. Students will analyze the commonalities and differences between the two documents, reflecting on how these might apply to artificial beings.

The activity begins with a screening of *Blade Runner 2049*, a film that raises ethical questions about the consciousness of automatons, their memories, and their emotional experiences. This sets the stage for an ethical discussion about the rights of non-human beings. After watching the film, students will engage in a debate centered on the rights of non-humans, with a specific focus on posthumanist theories. The discussion should explore the similarities and differences between the rights we currently recognize for humans and the potential rights for artificial beings.

Students will then move on to a research phase where they will examine in detail the components of the Universal Declaration of Human Rights and other similar legal frameworks. This comparative analysis will provide a foundation for understanding how a declaration of rights for artificial beings might be structured. Students will be divided into groups, with each group specializing in a specific section of the proposed declaration. Each group will be responsible for drafting their portion of the document, applying their research to propose appropriate rights and responsibilities for artificial beings. In the final phase, each group will present their proposed section of the declaration. A debate will follow, where students must defend their proposals, discuss potential amendments, and finally vote on each article. The process will culminate in a final version of the Declaration of Artificial Rights.

References

- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110, 5802 - 5805.
- Mitchell, M. (2019). *Artificial Intelligence: A Guide for Thinking Humans*. Farrar, Straus and Giroux.
- Müller, V. C. (2020). Ethics of artificial intelligence and robotics. In *The Stanford Encyclopedia of Philosophy*. Stanford University.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
- Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Alfred A. Knopf.

15

A Chalk with a Chip: Proposals for a 3.0 Classroom

New generative models have the potential to solve complex mathematical problems, design a curriculum for a new sociology course, adopt the persona of Kant or Hegel, compare two intellectual property laws and highlight key differences, simulate a conversation to resolve an ethical dilemma, transcribe an interview with digital nomads and code the results, identify significant contributions in the field of leisure psychology, participate in brainstorming sessions for a social marketing campaign, test students' knowledge in demography, clearly explain game theory and apply it to an energy control issue, translate a speech by the Chinese president about the New Silk Road from Cantonese, or suggest exercises to practice English prepositions. Artificial intelligence can enhance teaching to become more personalized, inclusive, international, multimodal, and engaging.

However, these new generative models can also create hallucinations in describing plate tectonics, introduce gender bias in sociological analysis, miss the ironic tone of a transcribed conversation, generate

predictable and uninspired knowledge of Hegelian dialectics, propose unoriginal ideas during brainstorming, dehumanize the tone of language, suggest exercises that reinforce already established concepts, lose track in a discussion on language theories, generate false information about labor theories, or fail to compare two legislative models accurately. AI has the potential to result in biased, dehumanized, monotonous, or even erroneous education.

We are at the dawn of a process whose future we cannot predict. We are like Stendhal at Waterloo, unable to foresee if AI will transform everything or, once again, if winter will come because the models fall short of expectations. This manual was written during this waiting period, in this limbo between what seems probable and what seems impossible. We anticipate that classrooms will change, but we cannot estimate how much or what the side effects will be. Based on limited experience and more intuition than conviction, we propose a provisional, transitional, and soon-to-be-outdated decalogue for the 3.0 classroom, awaiting either structural changes or another disappointment.

“We are unable to foresee if AI will transform everything or, once again, if winter will come because the models fall short of expectations.”

1. Advocating for Non-Technology

Technology is not essential in the classroom. Lessons can take place in a forest, amidst the noise of a busy street, in a simple room with nothing more than a chalkboard and some chalk, using salvaged slides, colorful post-its, or handwritten notes. Natural intelligence is a necessary and often the only viable option, because AI is dispensable, and an unplugged class is a legitimate one. The basements of educational institutions are filled with gadgets that were once heralded as revolutionary: digital boards, interactive screens, networked systems, virtual reality goggles, augmented reality, tablets, children’s robots, simulators...

In fact, classrooms need technology-free spaces, that is, detox areas where thought, speech, and writing take center stage. Just as industry hasn’t eliminated craftsmanship and kitchen robots don’t do barbecues, there are forms of knowledge transmission based on empathy, interaction, human contact, tone of voice, gestures, and silences that (for now) can only take place without technology. More than technological classrooms, digital spaces, 5G, and artificial intelligence, we need disconnected spaces and moments. In many cases, we need to disconnect in order to reconnect.

2. Reviving Cyberpunk

Artificial intelligence is a tool that can be used to amplify social tensions and exacerbate imbalances.

The primary educational task regarding AI is to prevent its misuse, fostering a critical and sometimes combative attitude against the abuses of a technology that conceals its processes in a black box. This requires a deep understanding of the inherent logic of neural networks, distinguishing between the two main schools of AI, evaluating the impact of reward systems and training processes, and considering indirect effects. Ultimately, it means reviving the cyberpunk mindset that anticipates potential dystopias in order to fight them. Indifference or ignorance will not allow for a critical alternative to unchecked artificial intelligence: educational institutions must become the operational grounds for resistance.

In the world of computing, there has been a long and persistent struggle in favor of free code, open-source software, accessible data, process transparency, networked collaboration, altruistic information exchange, open digital libraries, Wikipedia, hackathons, citizen science, online volunteering, and free educational platforms. This is the B-side of a model that prioritizes corporate profit over the cultural, social, and psychological impacts of technological tools. TikTok and Khan Academy are two outcomes of the same process: the digital revolution. Educational institutions must be one of the spaces where AI is shaped and modeled to serve our yet-to-be-realized utopias.

This also means fostering systems of information exchange: informal learning networks, open-

access documents, articles that share the results of teaching experiments, and meetings between educators who use these resources. This manual is part of an open-source, information-sharing mindset that can help create a neural network of teaching outcomes for the 3.0 classroom.

3. Purpose First, Tools Later

Carpenter tools are useful when we have a clear vision of the wardrobe we want to build. Too often, educational innovation has been an end in itself: doing things differently becomes the goal of change. In reality, AI has been much more beneficial when we clearly identify the gap we want to address. For example, in a master's program on cultural tourism, where some students have a solid background in tourism and others do not, the challenge arises of creating different content for two distinct groups (each with its own internal differences). This is a clear objective: how can common material be provided to two groups with different starting points? With AI support, the initial material can be adapted for students lacking prior knowledge, offering additional windows of content to help expand their understanding. First, we define the goal, then we choose the tools.

AI can solve very specific teaching problems: improving connection with a particular author, conveying a complex idea through a new narrative, developing dialectical skills with a virtual sparring partner, creating supplementary material to address comprehension difficulties on a particular topic, enhancing the ability to verify the accuracy of a source, guiding students through creative processes, answering questions outside of class hours, and fostering post-class participation. AI is far less effective when its primary objective is merely its own implementation. Students don't perceive the added value of the technology and don't understand why another approach wasn't taken to achieve the same goal. As we discussed in the first chapter, "asking is an art," and any application of AI must be a response to a relevant question.

4. The First Island of an Archipelago

It is more likely that AI implementation will succeed through small, targeted experiments within a traditional model than through a disruptive, massive application. AI has a fundamental issue: it is unpredictable. Its probabilistic, rather than deterministic, nature makes it impossible to set clear instructions that will guarantee specific outcomes. Various prompts, personal variations in natural language, the input of context, or output precision can lead to very diverse and often unpredictable results. This is not like a Python session, an SPSS workshop, or an open database training, where we follow a precise sequence to reach a set goal. There are many possible paths, and, in this case, the pedagogical focus should be more on the process and the journey than on the final result. A second challenge that consistently arises in AI-driven educational sessions is the temptation for students to use AI as an alternative route to the intended practice. For example, when we created a "Miró persona" to test students' knowledge of the artist through questions asked by a fictional journalist, some students used AI to generate the questions as well. The omnipresence and versatility of AI complicate the setting of boundaries between automated tasks and manual ones.

This is why the application of AI in the classroom requires a trial-and-error process, which is more feasible through small alternative practices within a traditional model than through a revolutionary overhaul. The risk of accumulating mistakes is high. Therefore, our teaching approach has always been to occupy small islands in an archipelago, creating routines that we can master after a few attempts, allowing us to consider further explorations into the open sea.

5. The Cyborg Student

It is very tempting to view AI as a substitute for intellectual work. A student tasked with writing an essay on the influence of Paul Klee's trip to Tunis or on the geopolitics of port control in the

Indian Ocean might easily opt to have AI do the work. In fact, there are thousands, if not tens of thousands, of papers in fields like philology, history, communication, geography, law, economics, sociology, marketing, or art that have been entirely written by a system like ChatGPT. It's challenging to convey to students that by relying solely on AI, they lose the opportunity to develop essential skills needed for writing: researching, cross-referencing information, selecting and reading contextual works, synthesizing ideas, and effectively communicating results. The only skill they develop in this case is their interaction with artificial intelligence.

We've already lost this battle when it comes to internet use. Since we haven't guided students on how to use the internet responsibly and efficiently, information searching is no longer seen as a complex skill based on evaluating sources, selecting, and prioritizing information. Instead, students often present a collage of content pulled from the web, where a blog post or data from a local magazine is given the same weight as an official source.

However, artificial intelligence can also be a controlled tool that enhances students' abilities. An efficient and critical use of AI can speed up information gathering, evaluate alternatives, improve contextual understanding, clarify vague concepts, compare documents, aid synthesis, establish non-obvious relationships, reinforce viewpoints, stimulate critical thinking, or improve the graphic quality of presentations. If AI accompanies a complex knowledge-building process, it has enormous potential to adapt, personalize, and expand the learning process. For certain tasks, it can act as a shortcut or accelerator, reducing time spent on some routines and allowing students to focus on other parts of the knowledge chain. In a way, the student becomes a cyborg—with enhanced or expanded skills but maintaining human control over the entire process. This is likely the great challenge for future 3.0 education: shifting from AI as a replacement to AI as a complement, from the absent student to the augmented student.

6. Responsible AI

Training OpenAI's GPT-3 model, which has 175 billion parameters, required an estimated 1.287 MWh of energy. This is equivalent to the CO₂ emissions of driving an average car for about 626,000 kilometers. The manufacturing of semiconductors requires large quantities of ultra-pure water; for instance, it's estimated that around 7,570 liters of water are needed to produce a single 30 cm chip. A typical data center can use up to 1.8 liters of water per kWh consumed to cool servers. Although companies rarely disclose the environmental impact of their processes, Google has reported that its data center in The Dalles, Oregon, used approximately 1.2 billion liters of water in 2021 for cooling. The extraction of materials necessary for semiconductor production, such as lithium, cobalt, and other rare earth elements, also has a significant environmental impact, including habitat destruction and water pollution.

This impact must be put into perspective. It is estimated that answering five questions on ChatGPT consumes about half a liter of water. In comparison, the water footprint of a beef steak is 7,000 liters, which means one steak is equivalent to 70,000 questions on ChatGPT. This doesn't imply that this technology is harmless, but it does highlight the complexity of environmental issues, which require diverse solutions. Responsible use of AI involves limiting questions and interactions to those that are truly necessary, avoiding repetitive processes that could be simplified with more specific interactions. The question we might ask ourselves is: "Do I really need to ask this question, or is there an equally valid alternative way to solve the problem?"

Responsible consumption also entails using models that offer higher environmental guarantees. For example, Anthropic's "Claude" is based on a stronger commitment to responsibility compared to other models with a clear commercial focus, such as ChatGPT, Gemini, or Copilot. Llama is also an interesting alternative because it uses open-source code, though its use is still restricted to Europe and can only be accessed via a VPN.

Artificial intelligence faces highly complex moral challenges, but also more specific moral decisions, such as those autonomous vehicles might encounter in high-risk situations. This ties into the trolley dilemma, a thought experiment in moral philosophy that tests our ethical intuitions. The dilemma was first introduced by British philosopher Philippa Foot in 1967, but it has since seen numerous variations. Imagine a runaway trolley speeding down a track. Further ahead on this track, there are five people tied up and unable to move. If the trolley continues on its path, it will kill the five people. However, an individual stands next to a lever that could divert the trolley onto a different track. On this secondary track, though, there is one person tied up, who will also die if the trolley is diverted. The individual can either do nothing, allowing the trolley to kill the five people on the main track, or pull the lever, saving the five but causing the death of the person on the secondary track.

This dilemma pits utilitarianism against deontology. Utilitarianism suggests that the morally right action is the one that generates the greatest good for the greatest number of people. From this perspective, diverting the trolley to save five lives at the cost of one would be the correct choice. Deontology, a philosophy associated with thinkers like Immanuel Kant, argues that certain actions are morally wrong regardless of their consequences. From this viewpoint, diverting the trolley could be seen as wrong because it involves actively causing the death of one person. The trolley dilemma remains a popular topic in philosophical discussions, as it raises tough questions about the nature of morality and our ethical intuitions. It also serves as a guide for understanding ethical debates related to AI.

7. Dual Training

There's a manual way to calculate standard deviation, and there are various programs that can do it automatically, allowing you to focus on interpreting the results. Which makes more sense: the manual method or the automated one? The

automatic system provides a standard deviation in seconds, enabling you to calculate many deviations, especially with large datasets, and compare results to evaluate which variables show more dispersion. However, learning the manual process helps you understand what the machine is doing. It also serves a second purpose: it allows you to calculate a standard deviation even if the program isn't working or there's a power outage. This dual ability to generate results both manually and automatically is crucial when working with new complex systems.

The growing development of increasingly precise tools will provide us with much greater capacities in areas like information searching, selection, comparison, analysis, argument development, theorization, and synthesis. However, training in skills without digital resources serves, first, to help us understand the processes operating inside AI's black box, and more importantly, it grants students a high degree of autonomy. One of the most significant risks of over-reliance on AI (even more than issues like privacy, bias, or fraud) is dependency. A price increase, a new access policy, a server crash, or a cyberattack could disrupt access to AI. The more extensive the training outside of AI, the lesser the impact of such scenarios.

8. Inside the Machine

Training students in AI-related skills also means introducing them to the understanding and manipulation of the machine itself. For example, it's highly recommended that students develop basic programming skills, such as learning the Python language. This multiplies their ability to connect environments and processes, linking AI's power with external tools, and improves their understanding of the routines that generate language models. In ChatGPT version 4, for instance, the system frequently shows Python code for the decisions it's making, which helps users interpret the operations and guide the model in a particular direction.

Having knowledge in object-oriented programming

and the design of algorithms and data structures should be a transversal skill. A deeper understanding of deep learning may require more advanced training, but students should at least have an intuitive notion of what an artificial neural network is and its various types. Presenting classic AI problems (such as Mary's Room) or new experiments, like [Claude's Golden Gate](#), can provide a more profound understanding of the opaque mechanisms that operate in AI systems.

9. Programmed Serendipity

In an increasingly algorithmic world, where AI provides precise and personalized answers, we risk losing creative randomness—those serendipitous encounters with unexpected ideas that often lead to great leaps in knowledge. The 3.0 classroom must preserve and encourage these spaces for serendipity, but in a conscious and structured way. We can program AI to surprise us, to introduce random elements into its responses, and to push us out of our intellectual comfort zones.

For example, we could design a “chaos bot” that, in the middle of a class on circular economy, introduces concepts from Norse mythology, forcing students to make improbable connections. Or a “forced analogy generator” that links, say, game theory with the art of bonsai. These exercises not only stimulate creativity but also train the mind to see patterns and relationships where there appear to be none—an increasingly valuable skill in a complex, interconnected world. Programmed serendipity is not a contradiction; it's a way to use AI to reclaim what AI itself threatens to eliminate: surprise, productive wandering, and lateral thinking. In the 3.0 classroom, the unexpected should be an integral part of the curriculum, not an anomaly.

10. Ethics as the Operating System

If AI is the hardware of the 3.0 classroom, ethics must be its operating system. It's not just about teaching ethics as another subject but integrating it into every interaction with AI, every decision about how to use it, and every reflection on its implications. The cyborg student we discussed earlier should not only be more efficient but also more conscious and responsible.

Imagine an “ethical assistant” that accompanies every AI query, raising moral dilemmas related to each topic, prompting reflection on the consequences of every decision. For example, in a marketing class, every AI-suggested strategy could come with an analysis of its potential social and environmental impacts. Or in a law class, we could program AI to automatically generate dystopian scenarios based on each new law studied, sparking debates about its potential authoritarian consequences. The goal is to create an artificial conscience—not in the sense of self-aware AI, but rather a system that constantly reminds us of our ethical responsibilities when using technology. In the 3.0 classroom, ethics shouldn't be an add-on but the framework within which all learning takes place. Only then can we form professionals who not only know how to use AI but also know how to question it, limit it, and, ultimately, humanize it.

José Antonio Donaire

Mònica Puntí

Konstantina Zerva

Raquel Camprubí

Núria Galí

From chalk to chip

The use of Artificial
Intelligence in the
Classroom

 Universitat
de Girona

ISBN: 84-8458-713-4



9 788484 587132