# Digital Skills for Information Use

## An interactive guide · Bachelor's degree (Transversal Training)

Pablo Sánchez-Núñez
Universidad Carlos III de Madrid

26 November 2026

## Contents

Welcome ..... 1
Using R in this book ..... 4
Unit 1. Introduction. Information sources ..... 6
Unit 2. Information retrieval ..... 18
Unit 3. Ethical use of information and plagiarism ..... 28
Units 4 and 5. References ..... 38
Units 6 and 7. Generative artificial intelligence: foundations and applications ..... 46
Units 8 and 9. Generative artificial intelligence: ethics, authorship and regulation ..... 55
Quick guide to APA 7 citations ..... 63
Checklists ..... 65
Verification lab ..... 66
Guide for teaching staff ..... 70
Glossary ..... 72
References ..... 75
Bibliography ..... 75

## Welcome

ℹ About this book

- **Author:** Pablo Sánchez-Núñez
- **ORCID:** https://orcid.org/0000-0001-7845-9506
- **Author website:** https://pablosancheznunez.com
- **Affiliation:** Universidad Carlos III de Madrid
- **DOI:** https://doi.org/10.5281/zenodo.24550050
- **Licence:** CC BY 4.0
- **Interactive web version:** https://pablosancheznunez.com/digital-skills-information/
- **How to cite:** Sánchez-Núñez, P. (2026). *Digital Skills for Information Use: an interactive guide*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24550050

ℹ About this PDF version

This is the version for reading and printing. Self-assessment questions are shown without automatic marking and the interactive activities appear on paper with their solutions; the labs only work on the web. To do the exercises with feedback and answer the questions, use the web version: https://pablosancheznunez.com/digital-skills-information/.

This book accompanies the course **Digital Skills for Information Use** (*Competencias digitales para el uso de la información*, code 19408, 3 ECTS, first year, second term), part of the Bachelor's degree **Transversal Training** programme (*Formación Transversal*) at Universidad Carlos III de Madrid. Its aim is to help you **find, evaluate and use information reliably and ethically**, including when it is produced by artificial intelligence.

The book follows the four blocks of the official syllabus: evaluating the credibility of digital information, searching for and organising information, using it ethically (including references and plagiarism) and generative artificial intelligence for content creation.

This is the English edition of the book. The Spanish edition has the same content.

**Author:** Pablo Sánchez-Núñez, Universidad Carlos III de Madrid
**ORCID:** 0000-0001-7845-9506
**Website:** pablosancheznunez.com

**DOI:** 10.5281/zenodo.24550050.

## Course objectives

By the end of the course you should be able to:

- Recognise information overload and disinformation, and use digital information critically.
- Identify, select and evaluate information sources, including social media as a source of scientific information.
- Design search strategies on general and specialised platforms, and organise what you find.
- Use information ethically, respecting intellectual property and avoiding plagiarism.
- Create and manage references with a citation style.
- Understand how generative artificial intelligence works, what it is useful for, what its limits are and what ethics and regulation say about its use.

## How each chapter is organised

Each chapter begins with **objectives** and a **glossary**, suggests a **required reading** and further readings, develops the **contents** of the syllabus, proposes a **practical exercise** and an **assessed challenge**, and ends with a **self-assessment** with instant feedback and **further resources**.

> **A little R**
>
> Four chapters include very short R exercises that run in your browser without installing anything, so you can look inside some tools: the Boolean operators of a search engine, the format of a reference, detecting overlaps between texts or checking references generated by an AI. You do not need to know how to program: the chapter Using R in this book explains the essentials, and every exercise has a hint and a solution.

> **Download the book**
>
> Book (PDF) Book (EPUB)
>
> The PDF is for reading or printing; the EPUB is for e-readers and phones.

## Your progress and offline use

The book records the exercises you solve and the questions you answer correctly. Your progress is stored **only in this browser**: it is not sent to any server and teaching staff cannot see it.

Turn on JavaScript to see your progress.

After your first visit, the browser keeps a copy of the book and the R engine, so you can read the chapters you have opened **offline**. If your browser supports it, this is enabled when the page loads.

Use the switch next to the book title, in the sidebar, to change between light and dark mode.

## Book map

| Block | Chapter | Syllabus units |
|---|---|---|
| Credibility of information | Information sources | 1 |
| Searching and organising | Information retrieval | 2 |
| Ethical use | Ethical use of information and plagiarism, References | 3, 4 and 5 |
| Generative AI | Foundations and applications, Ethics, authorship and regulation | 6 to 9 |

## Assessment

The course is taught in person with **continuous assessment (100 %)**: practical assignments and presentations (**70 %**) and a **final test** on the last day of class (**30 %**), in which you need at least 4 out of 10. Attendance and participation may count. Plagiarism or copying in any activity means losing the mark for that activity.

- Each chapter's **assessed challenges** are designed as practical assignments.
- The **self-assessments** do not count towards your mark and prepare you for the final test; they are also available in Aula Global.

This summary follows the official course description; if there is any difference, what the course publishes in Aula Global prevails.

## Core bibliography


- Bobish, G.; Jacobson, T. (eds.) (2014). *The information literacy user's guide: An open, online textbook*. Open SUNY Textbooks.
- Cordón García, J. A. et al. (2016). *Las nuevas fuentes de información: la búsqueda informativa, documental y de investigación en el ámbito digital*. Pirámide.
- Cruz Gil, C. (2015). *Modelos de búsqueda y recuperación de la información*. Trea.
- Dorner, D. G.; Gorman, G. E.; Calvert, P. J. (2015). *Information needs analysis: Principles and practice in information organizations*. Facet.
- Pacios Lozano, A. R. (coord.) (2013). *Técnicas de búsqueda y uso de la información*. Ramón Areces.


## Authorship and how to cite


Written by Pablo Sánchez-Núñez (Universidad Carlos III de Madrid).


Texts and figures are released under a CC BY 4.0 licence and the code under the MIT licence. The source code is on GitHub.



> Sánchez-Núñez, P. (2026). *Digital Skills for Information Use: an interactive guide*. Universidad Carlos III de Madrid. Zenodo. https://doi.org/10.5281/zenodo.24550050

DOI for all versions: 10.5281/zenodo.24550050; it always points to the latest edition. Each edition also has its own DOI, listed on the Zenodo record.

# Using R in this book

Some chapters have short R exercises that run **in your browser**. You are not going to learn to program: you will use R as a powerful calculator to look inside some information tools. This chapter explains just what you need in ten minutes.

> **Objectives**
>
> - Run and modify R code in the browser.
> - Store values and text in objects and work with vectors.

## How the code cells work

Each code block has a **Run Code** button. You can change the code and run it again; **Start Over** restores the original version. The first time you open a page, the browser downloads the R engine (a few seconds).

```
results <- 1250          # documents returned by a search
relevant <- 37           # the ones that are actually useful
relevant / results       # proportion
```

`<-` stores a value in an object. A **vector** groups several values with `c()`. Text goes in quotation marks:

```
sources <- c("Dialnet", "Google Scholar", "Scopus", "Wikipedia")
length(sources)                  # how many there are
"Scopus" %in% sources            # is Scopus there?
toupper(sources)                 # in capitals
```

### Exercise 1

A search returns 480 results and, after reviewing them, 24 are relevant. Calculate the proportion of relevant results (the **precision** of the search).

> **Exercise: fill in the blanks** _____
>
> ```
> relevant <- 24
> results <- ______
> relevant / results
> ```

> **Hint**
>
> Replace the blank with the total number of results.

> **Solution**
>
> ```
> relevant <- 24
> results <- 480
> relevant / results
> ```
>
> 0.05: only one result in twenty is useful. In the chapter on retrieval you will see how to improve precision.

### Exercise 2

Check whether “Dialnet” is in the vector `sources`.

**Exercise: fill in the blanks _____**

```
"_______" %in% sources
```

**Hint**

Write the name exactly as it appears, with a capital first letter.

**Solution**

```
"Dialnet" %in% sources
```

`TRUE`. If you write “dialnet” in lower case, R says `FALSE`: it is case-sensitive, like some search engines.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. In R, text is written in quotation marks. *[ T / F ]*
2. To create a vector with several values you use the function *[ length() / c() / toupper() ]*
3. `length(c("a", "b", "c"))` returns _____
4. To do the exercises in this book you need to install R on your computer. *[ T / F ]*

**Explanations**

- Without quotation marks, R would look for an object with that name.
- `c()` combines values into a vector.
- Three elements.
- They run in the browser with webR.

# Unit 1. Introduction. Information sources

**Objectives**

- Recognise information overload and the different types of false information.
- Use digital information critically, reflectively and proactively.
- Identify, select and evaluate information sources, including social media as a scientific source.

**Unit glossary**

Information overload · Disinformation · Misinformation · Credibility · Lateral reading · SIFT method · Primary source · Peer review · Altmetrics

**Downloads**

Challenge submission template (.qmd) Style theme (.scss)

Use the `.qmd` template and its `.scss` theme to write up your challenges: save them in the same folder.

## Required reading

Metzger, M. J.; Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210-220.

### Further reading

- Bawden & Robinson (2009) — Information overload and its effects.
- Pennycook & Rand (2021) — Why we believe and share fake news.
- Sugimoto et al. (2017)

## 1. Information, overload and false information

Accessing information has never been so easy, and knowing which information is reliable has never been so hard. Bawden & Robinson (2009) describe the "dark side" of information: **information overload** (more information than can be processed), the anxiety it causes and other paradoxical effects, such as feeling less informed the more information you receive.

When it comes to false information, it is useful to distinguish:

| Type | What it is | Example |
|---|---|---|
| **Misinformation** | False, but shared without intending to deceive | Forwarding an ineffective home remedy believing it works |
| **Disinformation** | False and created or spread to deceive | A campaign with invented data to influence an election |
| **Malinformation** | True, but used to cause harm | Publishing someone's real personal data |

Pennycook & Rand (2021) show that many people share false information not so much because of ideological bias as because of **inattention**: they do not stop to think whether it is true before sharing it. A short pause improves judgement.

## 2. Critical, reflective and proactive use

Metzger & Flanagin (2013) explain that, online, people do not assess credibility systematically but with **mental shortcuts** (heuristics): **reputation** (a well-known brand seems reliable), **endorsement** by others (many reviews or likes), **consistency** (what is repeated on several sites), **expectancy violation** (a sloppy design or errors make us suspicious) and **persuasive intent** (if it looks like they want to sell us something, we distrust it). These shortcuts are useful, but they can be exploited: a fake website can imitate a well-known brand and reviews can be bought.

Using information critically means replacing the shortcut with a check when it matters. For this, **lateral reading** works better than vertical reading: instead of analysing the page in depth (it may be designed to look reliable), you open other tabs to find out what **others** say about that source and that claim. This is what professional fact-checkers do.

The **SIFT** method, proposed by Mike Caulfield, sums up lateral reading in four steps:

1. **Stop**: before reading or sharing, ask yourself whether you know the source.
2. **Investigate the source**: who is behind it and what interests do they have?
3. **Find better coverage**: what do other reliable sources say about the same thing?
4. **Trace claims**: find the primary source of the claim, quote or image.

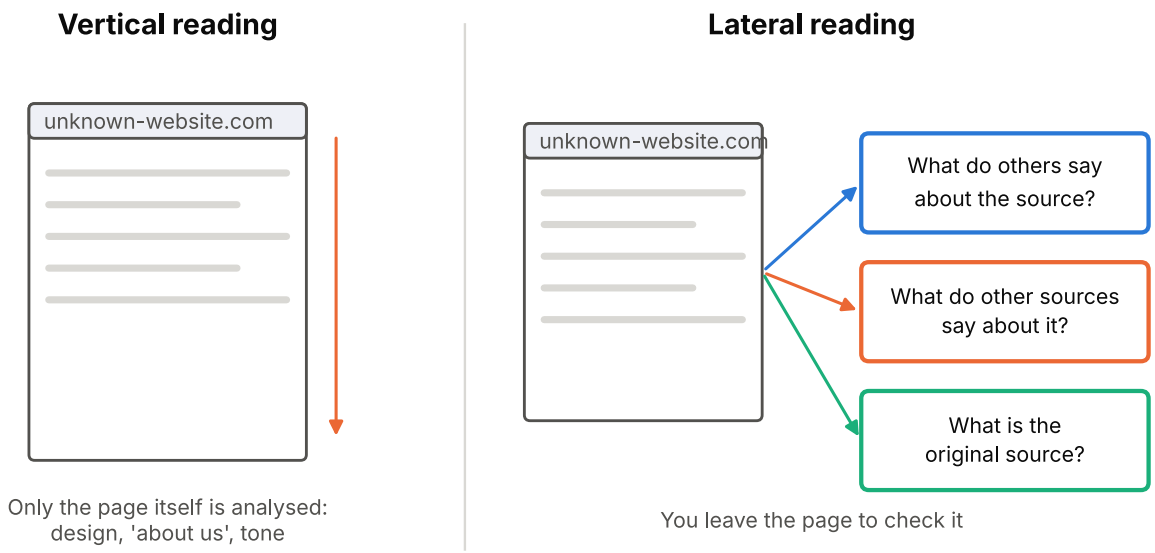


Figure 1: Vertical versus lateral reading. Lateral reading leaves the page to check the source and the claim in other sources.

## 3. Identifying, selecting and evaluating sources

Sources are usually classified as **primary** (original information: a research article, a dataset, a law, an interview), **secondary** (they analyse or summarise primary sources: a review, a textbook, a news story about a study) and **tertiary** (they help you find the others: catalogues, databases, encyclopaedias).

To select them, the key question is **what do I need**: a definition, up-to-date data, the scientific evidence on a topic, the legislation in force. To evaluate them, a useful checklist is:

| Criterion | Questions |
|---|---|
| **Authorship** | Who signs it? Do they know the subject? Is there a responsible organisation? |
| **Purpose** | Does it inform, give an opinion, sell, persuade? |

| Criterion | Questions |
|---|---|
| **Currency** | When was it published or updated? Does the date matter for your topic? |
| **Accuracy** | Does it cite its sources? Can it be checked? Does it agree with other reliable sources? |
| **Review** | Has it passed peer review, professional editing or no check at all? |

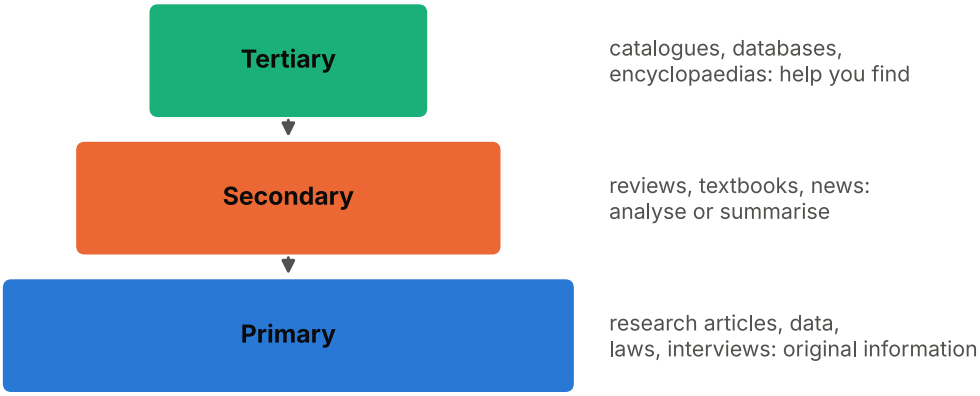


Figure 2: Types of sources. We often find a tertiary or secondary source first; the job is to get to the primary one.

## 4. Social media as a source of scientific information

Social media is increasingly used to spread and discuss science. Sugimoto et al. (2017) review how researchers use it and how these mentions are measured (**altmetrics**). For a student, social media can help you **discover** recent research and debates, but it is not the source: you need to get to the original article. Remember that:

- A viral post about a study may exaggerate or misinterpret its results.
- Social media mentions measure **attention**, not quality or truth.
- Preprints (versions before peer review) spread widely on social media and may change or never be published.

## 5. Disinformation and libraries: the IFLA infographic

In 2017 the International Federation of Library Associations and Institutions (IFLA) published the infographic *How to Spot Fake News*, adapted from a 2016 FactCheck.org article. It starts from one idea: **critical thinking** is a key skill of **media and information literacy**, and libraries should teach and defend it. IFLA offers it for download, printing, translation and sharing in libraries, communities and social media.

| Step | What the infographic proposes | In practice |
|---|---|---|
| **1. Consider the source** | Investigate the website, its mission and its contact information | Who publishes it? Is there a real "About us" and contact? |
| **2. Read beyond** | A striking headline seeks attention: what is the whole story? | Read the whole text, not just the headline or thumbnail |
| **3. Check the author** | A quick search: are they credible? Are they real? | Look for their professional profile or ORCID |

| Step | What the infographic proposes | In practice |
|---|---|---|
| **4. Supporting sources** | Click the links and check that the information supports the story | Trace the claim back to the original study or document |
| **5. Check the date** | Reposting old news does not make it relevant to current events | Look at the date of publication and of the data |
| **6. Is it a joke?** | If it is too outlandish, it may be satire | Check whether the outlet is satirical |
| **7. Check your biases** | Your own beliefs can affect your judgement | Be more suspicious of what confirms what you already thought |
| **8. Ask the experts** | Ask a librarian or consult a fact-checking site | University libraries, fact-checkers such as Full Fact or Maldita.es |

IFLA's eight steps and SIFT's four moves (worked example 1) complement each other: SIFT organises the process into four quick moves based on lateral reading, and IFLA breaks down what to check in each source. Step 7, **checking your own biases**, is perhaps the hardest: we tend to accept without checking whatever matches our ideas (confirmation bias).

Downloadable resource

IFLA (2017). *How to Spot Fake News* [Infographic]. International Federation of Library Associations and Institutions. https://repository.ifla.org/handle/20.500.14598/195 (available in English and in many other languages)

## Worked examples

**Example 1 · A news story about a study: lateral reading with SIFT**

A headline says: "Study proves that sleeping less than six hours doubles the risk of failing". The **SIFT** method (Mike Caulfield) proposes four quick moves:

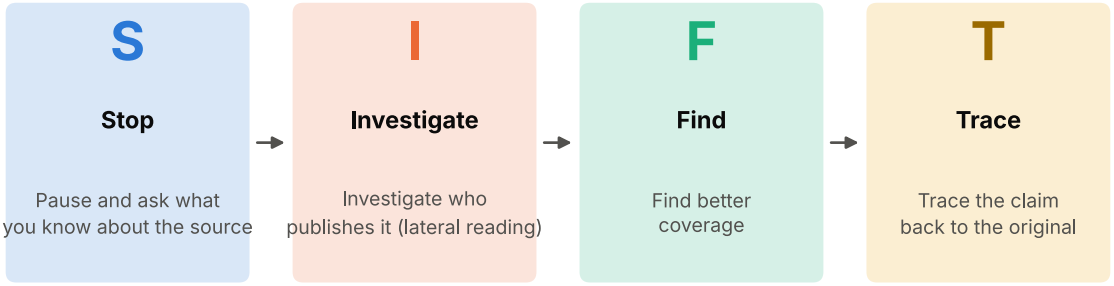


Figure 3: The four moves of the SIFT method for evaluating online information.

1. **Stop**: before sharing or believing it, ask yourself what you know about the source and how the headline makes you feel. A bold, emotional headline calls for more care.
2. **Investigate the source**: open another tab and find out who publishes the story (an outlet with a science desk, a commercial blog?). This is **lateral reading**: you evaluate a source by reading what others say about it, not what it says about itself.
3. **Find better coverage**: look for the topic in other reliable sources; if it is important, other outlets or agencies will have reported it.
4. **Trace claims**: find the original study (title, journal, DOI) and compare: is it an observational study? With how many people? Does it say "doubles" or "is associated with"?

**Typical result:** the original study found an **association** in a survey of students at one university; the headline turned it into a general causal relationship. The correct claim would be: "In a survey, students who slept less than six hours failed more subjects".

**Example 2 · Is this journal reliable?**

You receive an email inviting you to publish your bachelor's thesis in an "international, indexed journal with review in 72 hours" for €150. Checks, in order:

1. **Is it in recognised directories?** If it is open access, look for it in DOAJ (Directory of Open Access Journals). Check its ISSN on the ISSN portal.
2. **Where does it say it is indexed?** "Indexed" means nothing unless it says where; check in the databases themselves (Scopus, Web of Science, Dialnet, Latindex).
3. **Editorial board**: do these people exist? Do they know they are on the board? Are they from the field?
4. **Review**: serious peer review is not done in 72 hours.
5. **Checklist**: the *Think. Check. Submit.* initiative sums up these questions.

**Conclusion:** the combination of an unsolicited invitation, promised speed, vague indexing and a fee suggests a **predatory journal**. Do not send your work.

Example 3 · A viral thread about a preprint

A thread on social media announces a "discovery" and links to medRxiv. What you need to know: a **preprint** is work that has **not yet been peer reviewed**. It can be valuable and fast, but it should be read with caution: check whether it has already been published in a journal (the preprint page usually links to it), whether other specialists have commented on it and whether the thread exaggerates the results. When you cite it, always say it is a preprint.

## Graded exercises

**Basic.** Classify these sources as primary, secondary or tertiary: an article presenting an experiment; an encyclopaedia; a systematic review.

Explanations

Experiment: **primary** (it presents new data). Systematic review: **secondary** (it synthesises primary studies). Encyclopaedia: **tertiary** (it summarises and points to other sources).

**Intermediate.** Apply the four SIFT moves to a science news story from this week and write five lines on what changes between the headline and the original study.

Explanations

Check at least: who publishes it, whether there is other coverage, the design of the original study (experimental or observational), the sample size and whether the verb in the headline ("proves", "causes") matches what the study says ("is associated with", "suggests").

**Advanced.** Choose a journal in your field that you do not know and write a short report on its reliability using the five criteria in example 2 and the sources you consulted.

Explanations

A good report cites verifiable evidence (a link to DOAJ or the database, editorial board names that have been checked, the published review and fee policy) and distinguishes between what has been checked and what could not be verified.

## Activities

Activity: Which step of the IFLA infographic would you apply first?

**Categorías:** Check the date · Is it a joke? · Check your biases · Supporting sources · Check the author

- A 2019 story about a border closure is circulating today as if it were new
- An incredible headline on a website whose slogan is "news that never happened"
- An article that confirms exactly what you have always thought about a controversial topic
- "A study proves…" with no link or reference to the study
- A column signed by an "expert" with no professional trace you can find

Solution

- A 2019 story about a border closure is circulating today as if it were new → **Check the date**
- An incredible headline on a website whose slogan is "news that never happened" → **Is it a joke?**
- An article that confirms exactly what you have always thought about a controversial topic → **Check your biases**
- "A study proves…" with no link or reference to the study → **Supporting sources**
- A column signed by an "expert" with no professional trace you can find → **Check the author**

Activity: What type of false information is it?

Decide whether each case is misinformation, disinformation or malinformation.

**Categorías:** Misinformation · Disinformation · Malinformation

- Your aunt forwards a flu remedy on WhatsApp that she believes works
- A network of fake accounts spreads a doctored video of a candidate the night before an election
- Someone publishes a journalist's real address so that she is harassed
- A newspaper mistakenly publishes a wrong unemployment figure and corrects it the next day
- A website imitates a newspaper's design to spread an invented story
- Authentic private emails of a scientist are leaked, taken out of context, to discredit him

Solution

- Your aunt forwards a flu remedy on WhatsApp that she believes works → **Misinformation**. It is false, but she shares it without intending to deceive.
- A network of fake accounts spreads a doctored video of a candidate the night before an election → **Disinformation**. False and created to deceive.
- Someone publishes a journalist's real address so that she is harassed → **Malinformation**. It is true, but used to cause harm.
- A newspaper mistakenly publishes a wrong unemployment figure and corrects it the next day → **Misinformation**. An unintentional error; the correction is good practice.
- A website imitates a newspaper's design to spread an invented story → **Disinformation**. It impersonates a well-known brand to deceive.
- Authentic private emails of a scientist are leaked, taken out of context, to discredit him → **Malinformation**. Real material used to harm.

**Activity: Which mental shortcut is at work?**

Metzger & Flanagin (2013) describe five credibility heuristics. Identify which one is at work in each situation.

**Categorías:** Reputation · Endorsement · Consistency · Expectancy violation · Persuasive intent

- You trust an article because it appears on the website of a well-known university
- You choose a hotel because it has 4,000 five-star reviews
- You believe a figure because you have seen it on three different websites
- You distrust a website full of spelling mistakes and pop-up ads
- You doubt a study on sugar funded by a soft-drink manufacturer

**Solution**

- You trust an article because it appears on the website of a well-known university → **Reputation**
- You choose a hotel because it has 4,000 five-star reviews → **Endorsement**
- You believe a figure because you have seen it on three different websites → **Consistency**. Beware: three websites may copy the same source.
- You distrust a website full of spelling mistakes and pop-up ads → **Expectancy violation**
- You doubt a study on sugar funded by a soft-drink manufacturer → **Persuasive intent**

**Activity: Primary, secondary or tertiary**

**Categorías:** Primary · Secondary · Tertiary

- An article presenting the results of an experiment
- The dataset of a survey by the Spanish National Statistics Institute (INE)
- A systematic review on mobile phone use in class
- A news story summarising a study
- The text of the EU AI Act published in the Official Journal of the EU
- A university textbook
- The university library catalogue
- A bibliographic database such as Dialnet

Solution

- An article presenting the results of an experiment → **Primary**
- The dataset of a survey by the Spanish National Statistics Institute (INE) → **Primary**
- A systematic review on mobile phone use in class → **Secondary**
- A news story summarising a study → **Secondary**
- The text of the EU AI Act published in the Official Journal of the EU → **Primary**
- A university textbook → **Secondary**
- The university library catalogue → **Tertiary**
- A bibliographic database such as Dialnet → **Tertiary**

Activity: The SIFT steps

- Find better coverage in other reliable sources
- Investigate who is behind the source
- Stop before reading or sharing
- Trace the claim back to its origin

Solution

1. Stop before reading or sharing
2. Investigate who is behind the source
3. Find better coverage in other reliable sources
4. Trace the claim back to its origin

ℹ Scenario: A headline about coffee

A friend shares on social media: “Study proves three coffees a day cut the risk of Alzheimer’s by 60%”. It has 20,000 likes. (An invented case for practice.)

**1. You see it on your phone**

What do you do?

- I share it: it has lots of likes and I like coffee ⟶ step 2
- I stop and look at who published it ⟶ step 3

**2. Shared**

*(bad ending)*

You used the endorsement heuristic: likes measure popularity, not truth. And you have amplified something without checking it.

**3. Who published it?**

The post links to a website called “naturalhealthtoday.info” that you do not know.

- I read the website carefully: if it looks serious, I trust it ⟶ step 4
- I open another tab and search for what is said about that website ⟶ step 5

**4. Vertical reading**

*(could be better)*

The website has a good design and a very professional “About us” page. But design can be imitated: without leaving the page you do not know who is behind it. Try lateral reading.

**5. Lateral reading**

You discover that the website sells food supplements. You look for the headline in other sources.

- I look for the original study ⟶ step 7
- It is enough that other similar websites publish it ⟶ step 6

**6. Misleading consistency**

*(bad ending)*

Many websites copy the same press release: repetition is not verification. You still need to get to the original.

**7. The original study**

You find the article: it is an observational study that finds a modest **association** between coffee consumption and lower risk, and it warns that it does not prove causation. There is no “60%” for three coffees.

- I conclude that the headline exaggerates and I do not share it (or I share it with the context) ⟶ step 8
- I conclude that the study is fake ⟶ step 9

**8. Well-calibrated judgement**

*(good ending)*

You have made the four SIFT moves: stop, investigate the source, find better coverage and trace the original. The headline is **exaggerated**, not entirely false: it confuses association with cause and invents the figure.

**9. Too far**

*(could be better)*

The study exists and is legitimate; what fails is the headline. Distinguish between the original source and how it has been reported.

⚠ Common mistakes

- **Confusing popularity with reliability.** Likes, views and mentions measure attention.
- **Staying on the page.** A good design or a convincing "About us" proves nothing: leave the page.
- **Taking repetition as confirmation.** Ten websites copying the same press release are a single source.
- **Reading only the headline.** The headline summarises (and sometimes exaggerates) what the study says; read at least the original abstract.
- **Distrusting everything.** The goal is not total scepticism but calibrating your trust to the evidence.

## Practice

Choose a striking claim you have seen on social media this week. Apply the four SIFT steps and note how long it took you and what you found at each step. Compare it with what you would have concluded by reading only the post.

## Assessed challenge

! Unit 1 challenge · Verify a scientific claim

Choose a claim related to science or health that is circulating on social media or in the media:

1. Apply the **SIFT** method and document each step with screenshots and links.
2. Find the **primary source** (the original study, report or data) and compare what it says with what is claimed.
3. Evaluate **three sources** you used with the criteria table in section 3.
4. Conclude whether the claim is true, false, exaggerated or unverifiable, and explain why.

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student chooses a different claim and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. Forwarding an ineffective home remedy believing it works is a case of *[ disinformation / misinformation / malinformation ]*
2. According to Pennycook and Rand, many people share fake news mainly through a lack of _____
3. Lateral reading consists of analysing the page itself in depth before leaving it. *[ T / F ]*
4. The last SIFT step, tracing to the original, looks for the *[ tertiary / primary / secondary ]* source
5. Which heuristic do we use when we trust a website because it has many positive reviews? *[ Reputation / Endorsement / Persuasive intent ]*
6. A review article summarising many studies is a *[ primary / secondary / tertiary ]* source
7. What do social media mentions of an article measure?

*(a) The scientific quality of the article (b) The attention it receives (c) Whether its results are true*

8. Spreading someone's real personal data to harm them is *[ misinformation / disinformation / malinformation ]*
9. A law published in the Spanish Official State Gazette (BOE) is a *[ primary / secondary / tertiary ]* source
10. A preprint has already passed peer review. *[ T / F ]*

Explanations

- It is false, but without intent to deceive.
- A pause to think improves judgement.
- That is vertical reading; lateral reading leaves the page to check it.
- The original source of the claim.
- Endorsement by others: reviews, likes, recommendations.
- It summarises and analyses primary sources.
- Altmetrics measure attention, not quality or truth.
- It is true, but used to cause harm.
- It is the original text of the law.
- It is a version before review.

## Further resources

| Resource | What it is for |
|---|---|
| IFLA – How to Spot Fake News (2017) | Eight-step infographic for spotting fake news, in many languages |
| Caulfield, M. – SIFT (The four moves) | The SIFT method explained by its author |
| ACRL – Framework for Information Literacy | Framework for information literacy |
| Full Fact, Maldita.es and Newtral | Fact-checkers in the UK and Spain |
| IFCN – Code of principles | International fact-checking network |
| *The information literacy user's guide* | Open textbook from the core bibliography |

# Unit 2. Information retrieval

**Objectives**

- Apply the principles and strategies of information retrieval.
- Use general and specialised resource platforms.
- Organise the information you find.

**Unit glossary**

Information need · Boolean operator · Truncation · Field searching · Precision · Recall · Berrypicking · Controlled vocabulary

**Downloads**

Chapter R script Challenge submission template (.qmd) Style theme (.scss)

## Required reading

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, 13(5), 407-424.

### Further reading

- Marchionini (2006) — Searching as learning.
- Cruz Gil (2015) — Core course bibliography.
- Pacios Lozano (2013) — Core bibliography.

## 1. From the need to the strategy

Every search begins with an **information need**: what you want to know and why. Before typing anything into a search engine:

1. **Formulate the question** in one sentence: "How does mobile phone use affect university students' academic performance?".
2. **Identify the concepts**: mobile phone use, academic performance, university students.
3. **Find synonyms and related terms** for each concept, in English and, if useful, in other languages (*smartphone*, *mobile phone*, *cell phone*; *academic performance*, *achievement*; *university students*, *undergraduates*).
4. **Decide the limits**: language, dates, type of document.

Bates (1989) showed that real searching is not a single query that returns the perfect set, but a process she called ***berrypicking***: the need changes as you read, you pick up useful pieces at each step and you use a variety of techniques (following an article's references, searching for an author, browsing a journal). Marchionini (2006) calls **exploratory search** the kind that is used to learn about a topic, as opposed to looking up a single fact.

## 2. Operators and techniques

| Technique | What it does | Example |
|---|---|---|
| **AND** | All the terms: reduces results | `mobile AND performance` |
| **OR** | Any of the terms: broadens (synonyms) | `mobile OR smartphone` |
| **NOT** | Excludes a term | `performance NOT sports` |

| Technique | What it does | Example |
|---|---|---|
| **Parentheses** | Group and order the operations | `(mobile OR smartphone) AND performance` |
| **Quotation marks** | Exact phrase | `"academic performance"` |
| **Truncation** | Variants of a stem | `universit*` finds university, universities… |
| **Field searching** | Only in title, author, abstract… | `title:(smartphone)` depending on the platform |

Each platform has its own syntax; each search engine's help pages explain it. In addition, many databases use a **controlled vocabulary** (thesauri, subject headings): standardised terms that group synonyms and improve results.

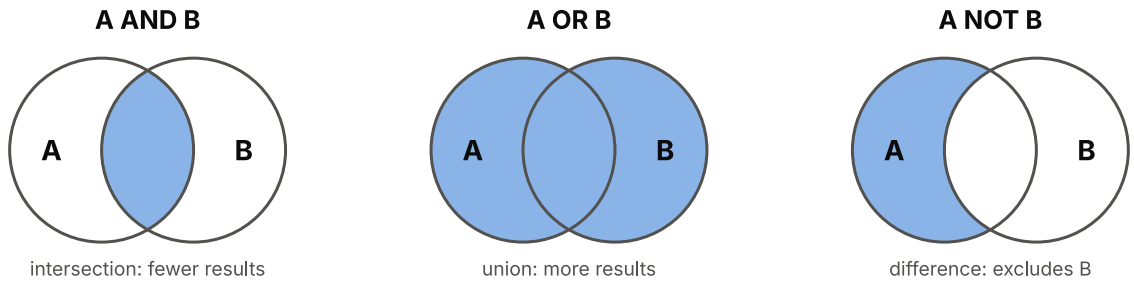


Figure 4: Boolean operators as sets. AND keeps the intersection; OR, the union; NOT, the difference.

## 3. Precision and recall

Two classic measures describe a search:

- **Precision**: of what I have found, what proportion is useful?
- **Recall**: of everything that would be useful, what proportion have I found?

They usually move in opposite directions: adding synonyms with OR increases recall and lowers precision; adding concepts with AND increases precision and lowers recall. A class assignment needs precision above all; a literature review needs recall.

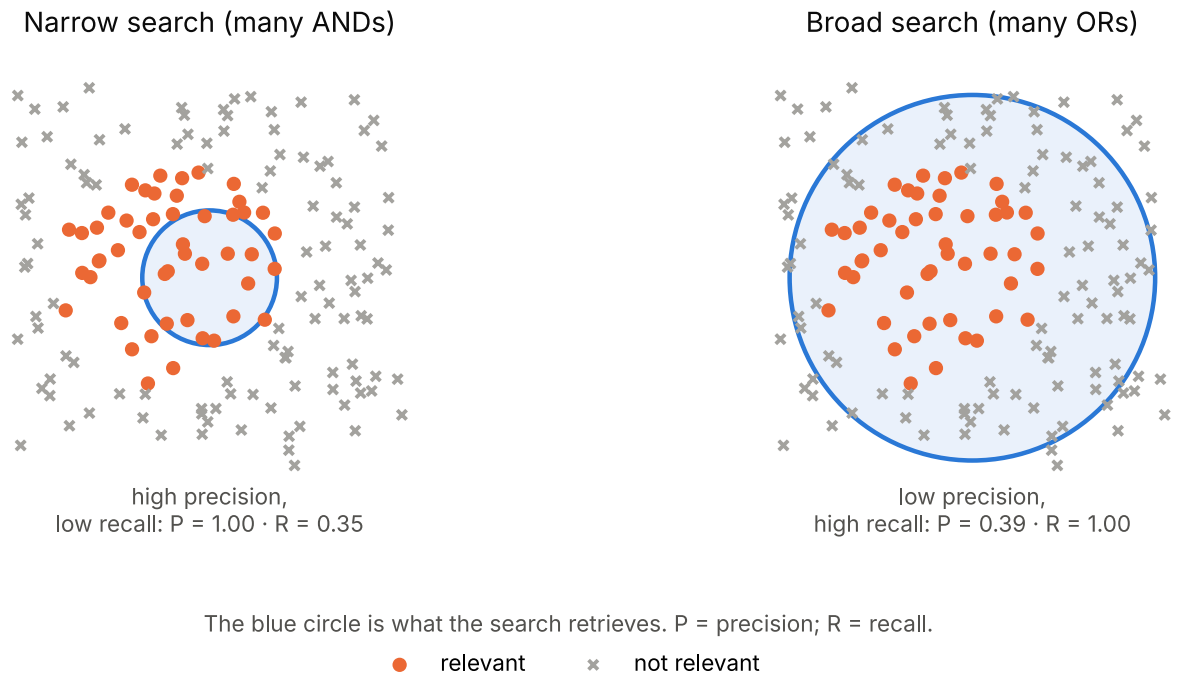


Figure 5: Precision and recall with the same documents. A narrow search retrieves few documents, almost all relevant; a broad one retrieves almost all of them, but with a lot of noise.

## 4. General and specialised platforms

| Type | Examples | What for |
|---|---|---|
| General search engines | Google, Bing | General information, institutional websites, news |
| Academic search engines | Google Scholar, BASE, OpenAlex | Scientific literature in all fields |
| Library catalogues | UC3M library catalogue | Books and resources subscribed to by the university |
| Multidisciplinary databases | Web of Science, Scopus (with university access), Dialnet | Peer-reviewed articles, with filters and controlled vocabulary; Dialnet for literature in Spanish |
| Specialised databases | PubMed (biomedicine), ERIC (education), LISTA (library science) | Comprehensive searching in one field |
| Repositories | Recolecta, institutional repositories | Open access versions |

Some resources require you to log in with your university account or use the VPN from outside the campus.

## 5. Organising information

Finding is half the job: the other half is not losing what you found. It helps to:

- **Save your search strategy** (where, with which terms, when and how many results), so you can repeat it and explain it.
- Use a **reference manager** (Unit 4) to save documents with their details.
- Set up **alerts** in databases to receive new results.
- Take notes with the full reference from the start, so you do not have to look for it later.

## R practice: Boolean operators from the inside

A search engine treats each term as the **set** of documents that contain it. In R, `intersect()` is AND, `union()` is OR and `setdiff()` is NOT:

```
mobile      <- c(1, 2, 3, 5, 8, 13, 21)     # documents containing "mobile"
smartphone  <- c(2, 4, 8, 16, 32)           # those containing "smartphone"
performance <- c(1, 2, 4, 8, 13, 16, 40)    # those containing "performance"

union(mobile, smartphone)                          # mobile OR smartphone
intersect(union(mobile, smartphone), performance)  # (mobile OR smartphone) AND
performance
```

### Exercise 2.1

How many documents does `(mobile OR smartphone) AND performance` return?

**i** Exercise: fill in the blanks _____

```
r <- intersect(union(mobile, smartphone), ______)
length(r)
```

**i** Hint

The third concept is `performance`.

Solution

```
r <- intersect(union(mobile, smartphone), performance)
length(r)
```

Six documents: 1, 2, 8, 13, 4 and 16. Without the synonym *smartphone* you would only have found four: adding synonyms with OR increases recall.

### Exercise 2.2

You have reviewed the six results and only four are useful. Calculate the **precision** of the search.

**i** Exercise: fill in the blanks _____

```
relevant <- 4
results <- 6
relevant / ______
```

**i** Hint

Precision is relevant divided by results.

Solution

```
relevant <- 4
results <- 6
relevant / results
```

0.67: two out of three results are useful.

## Lab: live Boolean operators

Interactive lab

Type two concepts (you can use OR, quotation marks and truncation within each one) and press **Search**. The book queries OpenAlex, an open index of more than 250 million scholarly works, and shows how many documents each combination retrieves. Watch how AND reduces, OR broadens and NOT excludes.

**Try:** add a synonym to A with OR and see how `A AND B` changes; then put a phrase in quotation marks and compare it with the same words without them.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Worked examples

Example 1 · From the question to the search string

**Question:** how does mobile phone use affect university students' academic performance?

1. **Split the question into concepts:** (a) mobile phone, (b) academic performance, (c) university students.
2. **Gather synonyms and variants** for each concept:

| Concept | Terms |
|---|---|
| Mobile phone | smartphone*, "mobile phone", *"cell phone"* |
| Academic performance | "academic performance", "academic achievement", grades |
| University students | universit, *"higher education", "college student"* |

3. **Combine with OR within each concept and with AND between concepts:**

```
(smartphone* OR "mobile phone*" OR "cell phone*") AND ("academic performance" OR "academic achievement" OR grades) AND (universit* OR "higher education" OR "college student*")
```

4. **Limit the fields** if there is too much noise: search in title, abstract and keywords (in Scopus, `TITLE-ABS-KEY(...)`).
5. **Document** the database, the date, the exact search string and the number of results.

Example 2 · Too many or too few results

- **12,000 results (low precision):** restrict to title and abstract; add a concept with AND; filter by year, language or document type; use exact phrases in quotation marks.
- **8 results (low recall):** add synonyms with OR; use truncation (`educa*` retrieves education, educational, educator…); remove the least essential concept; search more databases; check the references of relevant articles ("snowball" searching).

Example 3 · The same search on three platforms

| Platform | Strength | Watch out for |
|---|---|---|
| **Dialnet** | Literature in Spanish, journals and theses from Spanish universities | Less international coverage |
| **Scopus / Web of Science** | Selected international coverage, field searching, citations | Access through institutional subscription; less literature in Spanish |
| **Google Scholar** | Breadth, full text, free versions | Low precision, no reproducible search; limited operators (`intitle:`, quotation marks, -) |

It is best to combine them: start in a specialised or multidisciplinary database and complete with Google Scholar to find the full text.

## Graded exercises

**Basic.** What does `educa* AND (virtual OR online)` retrieve?

Explanations

Documents containing any word beginning with *educa* (education, educational, educator…) **and** at least one of the words *virtual* or *online*.

**Intermediate.** Write the search string for the question: "What effect does gamification have on secondary school students' motivation?".

Explanations

One possible solution: `(gamif* OR "serious game*" OR "game-based learning") AND (motivat* OR engagement) AND ("secondary school*" OR "secondary education" OR "high school*")`. Others are correct if they keep the structure: OR within each concept, AND between concepts.

**Advanced.** Run your search string in two databases, note the results, review the first 20 in each and calculate what proportion is relevant (approximate precision). Propose an adjustment to improve it.

Explanations

Present a table with database, date, search string, number of results and relevant documents among the first 20. The adjustment must be justified: if precision is low, restrict fields or add concepts; if known key documents are missing, add synonyms.

## Activities

Activity: From the need to the search

- Choose the platforms and adapt the syntax
- Combine the terms with operators
- Find synonyms and variants for each concept
- Formulate the question in one sentence
- Identify the main concepts
- Review the results and reformulate the search

Solution

1. Formulate the question in one sentence
2. Identify the main concepts
3. Find synonyms and variants for each concept
4. Combine the terms with operators
5. Choose the platforms and adapt the syntax
6. Review the results and reformulate the search

Activity: Where would you look first?

**Categorías:** Academic search engine · Specialised database · Library catalogue · Repository · General search engine

- A textbook from the core bibliography, to borrow it
- All published clinical trials on a drug
- The open access version of an article from a Spanish university
- A first look at the literature of any discipline
- The official website of a public body
- Articles on education with their own thesaurus

Solution

- A textbook from the core bibliography, to borrow it → **Library catalogue**
- All published clinical trials on a drug → **Specialised database**. PubMed, for example.
- The open access version of an article from a Spanish university → **Repository**
- A first look at the literature of any discipline → **Academic search engine**
- The official website of a public body → **General search engine**
- Articles on education with their own thesaurus → **Specialised database**. ERIC has its own thesaurus.

Activity: Which technique would you use?

**Categorías:** AND · OR · NOT · Quotation marks · Truncation

- Retrieve "library", "libraries" and "librarian" at the same time
- Add the synonym "smartphone" to "mobile"
- Require both "disinformation" and "vaccines" to appear
- Search for exactly "information literacy"
- Exclude results about sports performance

Solution

- Retrieve "library", "libraries" and "librarian" at the same time → **Truncation**
- Add the synonym "smartphone" to "mobile" → **OR**
- Require both "disinformation" and "vaccines" to appear → **AND**
- Search for exactly "information literacy" → **Quotation marks**
- Exclude results about sports performance → **NOT**

Common mistakes

- **Searching with a whole sentence** ("how does the mobile phone affect university students' performance"): the search engine does not understand the question; separate the concepts.
- **Forgetting other languages.** Most scientific literature is in English, but literature on local topics may be in Spanish or another language; search in both when relevant.
- **Sticking to the first page** of a search engine sorted by relevance without checking the filters.
- **Not saving the strategy.** If you do not note where, when and with which terms you searched, you will not be able to repeat or justify it.
- **Using NOT freely**: it also excludes useful documents that mention the term in passing.

## Assessed challenge

! Unit 2 challenge · A documented search strategy

Choose a topic you are interested in related to your degree:

1. Formulate the **question**, identify the concepts and make a **table of terms** with synonyms and variants.
2. Build the **search string** with operators, parentheses, quotation marks and truncation.
3. Run it on **three platforms** of different types (an academic search engine, a database and a catalogue or repository), adapting the syntax. Note the results for each one.
4. Select the **five** most relevant **documents**, explain why and calculate the precision of the first page of results on each platform.

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student chooses a different topic and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. To broaden a search with synonyms you use the operator *[ AND / OR / NOT ]*
2. `universit*` is an example of _____
3. Adding concepts with AND usually increases precision and reduces recall. *[ T / F ]*
4. According to Bates, real searching is a single query that returns the perfect set of documents. *[ T / F ]*
5. If I find 20 documents and 5 are useful, precision is *[ 0.20 / 0.25 / 0.50 / 4 ]*
6. What are quotation marks used for in a search?

*(a) To exclude a term (b) To search for an exact phrase (c) To search only in the title*

7. To search the biomedical literature comprehensively, the most suitable platform is *[ a general search engine / PubMed / an institutional repository ]*
8. The proportion of all relevant documents that a search retrieves is its *[ precision / recall / relevance ]*
9. A thesaurus is an example of a _____ vocabulary
10. In `(mobile OR smartphone) AND performance`, the parentheses make OR be resolved before AND. *[ T / F ]*

Explanations

- OR adds sets together: any of the terms.
- The asterisk retrieves all endings of the stem.
- Each AND requires one more concept: fewer, more relevant results.
- It is a berrypicking process that changes as you read.
- 5 / 20 = 0.25.
- Quotation marks fix the whole phrase, in order.
- PubMed is the specialised biomedical database.
- Recall is also called sensitivity.
- It standardises terms and groups synonyms.
- Just like in an arithmetic operation.

## Further resources

| Resource | What it is for |
|---|---|
| OpenAlex | Open index of the scientific literature |

| Resource | What it is for |
|---|---|
| BASE | Search engine for open access academic documents |
| Dialnet | Scientific literature in Spanish |
| Recolecta | Spanish open access repositories |
| *The information literacy user's guide*, search chapters | Open textbook |

# Unit 3. Ethical use of information and plagiarism

**Objectives**

- Know the basics of intellectual property and the right of quotation.
- Recognise the different types of plagiarism.
- Produce academic work without plagiarism: quote, paraphrase and summarise correctly.

**Unit glossary**

Intellectual property · Right of quotation · Public domain · Creative Commons licence · Plagiarism · Paraphrase · Patchwriting · Academic integrity

**Downloads**

Chapter R script Challenge submission template (.qmd) Style theme (.scss)

## Required reading

Pecorari, D.; Petrić, B. (2014). Plagiarism in second-language writing. *Language Teaching*, 47(3), 269-302.

**Further reading**

- Royal Legislative Decree 1/1996, Spanish Intellectual Property Act (in Spanish). Consolidated text; article 32 regulates quotations.
- Creative Commons ()

## 1. Information ethics

Using information ethically means acknowledging other people's work, respecting their rights and not deceiving the readers of your own work. At university, this commitment is called **academic integrity**: honesty, trust, fairness, respect and responsibility in study and research.

## 2. Intellectual property

In Spain, the **Intellectual Property Act** (Royal Legislative Decree 1/1996) protects original creations (texts, images, music, software…) from the moment they are created, without needing to register them. It distinguishes:

- **Moral rights**: recognition of authorship and integrity of the work. They cannot be waived.
- **Economic rights** (exploitation rights): reproduction, distribution, public communication and transformation. As a general rule, they last for the author's whole life and **70 years** after their death; after that the work enters the **public domain**.

The **right of quotation** (art. 32) allows you to include fragments of other people's works without asking permission when it is done as a quotation or for analysis, commentary or critical judgement, for teaching or research purposes, to the extent justified by that purpose, and **stating the source and the author's name**.

**Creative Commons licences** allow authors to authorise certain uses in advance. They combine four conditions: **BY** (attribution, present in all of them), **SA** (share alike), **NC** (non-commercial) and **ND** (no derivatives). This book, for example, is licensed CC BY 4.0.

| | Copy and distribute | Commercial use | Derivative works | Same licence |
|---|---|---|---|---|
| **CC BY** | ✓ yes | ✓ yes | ✓ yes | not required |
| **CC BY-SA** | ✓ yes | ✓ yes | ✓ yes | required |
| **CC BY-ND** | ✓ yes | ✓ yes | × no | not required |
| **CC BY-NC** | ✓ yes | × no | ✓ yes | not required |
| **CC BY-NC-SA** | ✓ yes | × no | ✓ yes | required |
| **CC BY-NC-ND** | ✓ yes | × no | × no | not required |

All require attribution (BY). SA: share alike · NC: non-commercial · ND: no derivatives.

Figure 6: What each Creative Commons licence allows.

## 3. What plagiarism is

**Plagiarism** is presenting someone else's work, ideas or words as your own without acknowledging them. Pecorari & Petrić (2014) distinguish **intentional** plagiarism from the kind that comes from not knowing how to write with sources, especially when learning to write in another language or in a new academic genre. Both are plagiarism, but the second is prevented by learning.

| Type | What it is |
|---|---|
| **Verbatim copying** | Copying text without quotation marks or citation |
| **Patchwriting** | Copying while changing a few words or their order, keeping the structure of the original |
| **Uncited paraphrase** | Rewording someone else's ideas well, but without citing the source |
| **Plagiarism of ideas** | Using someone else's argument, data or structure without acknowledging it |
| **Self-plagiarism** | Reusing your own work that has already been assessed or published without saying so |
| **Work done by others** | Submitting work commissioned from someone else or generated by an AI as if it were your own |

In this course, plagiarism or copying in any activity means **losing the mark for that activity**.

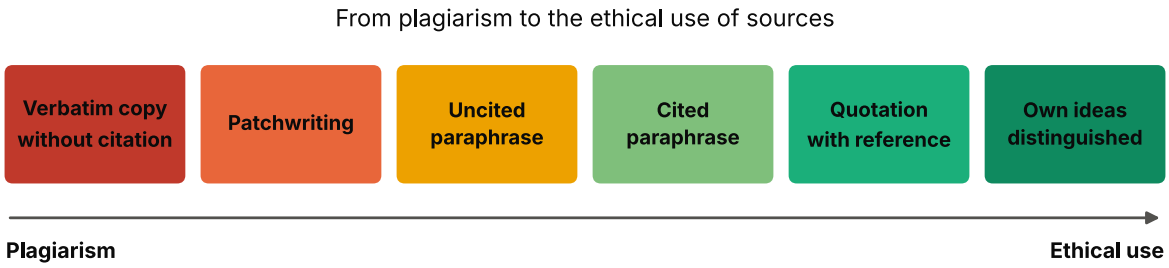


Figure 7: From plagiarism to ethical use: what changes is not the words, but whether you acknowledge the source.

## 4. How to write with sources without plagiarising

- **Quotation**: the exact words, in quotation marks, with the reference (and the page, if there is one). For sentences that matter because of their exact wording.
- **Paraphrase**: the author's idea in **your own words and your own structure**, with the reference. Tip: read, close the text, write what you have understood and then check that it is not too similar to the original.
- **Summary**: the main ideas of a long text, more briefly, with the reference.
- **Your own ideas**: they do not need a citation, but they must be clearly distinguished from other people's.
- **Common knowledge** (widely known, easily checked facts): no citation needed. When in doubt, cite.

## R practice: how software detects overlaps

Similarity-checking software compares fragments of consecutive words. A simple approach is to split texts into groups of three words (**trigrams**) and calculate what proportion they share (**Jaccard similarity**: shared / total distinct):

```
trigrams <- function(text) {
  p <- strsplit(tolower(gsub("[[:punct:]]", "", text)), "\\s+")[[1]]
  unique(paste(p[-c(length(p) - 1, length(p))], p[-c(1, length(p))], p[-(1:2)]))
}
similarity <- function(a, b) {
  ta <- trigrams(a); tb <- trigrams(b)
  length(intersect(ta, tb)) / length(union(ta, tb))
}
original    <- "Lateral reading means leaving the page to check what other sources say
about it"
patchwork   <- "Lateral reading means leaving the website to check what other sources
say about it"
paraphrase  <- "To verify a website it is better to consult what different sources think
of it than to analyse it from the inside"
c(patchwork = similarity(original, patchwork), paraphrase = similarity(original,
paraphrase))
```

### Exercise 3.1

Calculate the similarity between the original and a **verbatim copy** (the same text). What value do you expect?

ℹ Exercise: fill in the blanks _____

```
copy <- original
similarity(original, ______)
```

Hint

Compare the original with the object `copy`.

Solution

```
copy <- original
similarity(original, copy)
```

1: all the trigrams match. Patchwriting gives an intermediate value and a good paraphrase, almost 0. That is why this software detects copying well, patchwriting less well and paraphrase or translation hardly at all: **it does not replace the judgement** of the marker.

## Lab: what a similarity checker sees

Interactive lab

Paste an original text and your version. The lab splits both into groups of three words (**trigrams**), calculates what proportion they share (Jaccard similarity) and **highlights** the words in your version that belong to copied trigrams. Try a verbatim copy, a patchwriting and a real paraphrase.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Worked examples

**i** Example 1 · Paraphrasing step by step

**Original text** (invented for this example): "Peer review, despite its limitations, remains the main quality-control mechanism of the scientific literature, although its effectiveness in detecting errors has been questioned by numerous studies" (Pérez, 2021, p. 14).

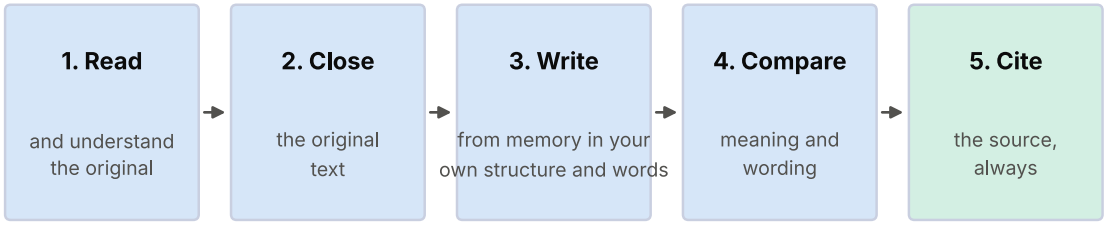


Figure 8: Five steps to paraphrase without plagiarising.

1. **Read and understand** the whole text; then **close** the original.
2. **Write from memory** the idea in your own words and structure: *Although it is debated whether it catches errors well, evaluation by specialists continues to be the central tool for guaranteeing the quality of what is published in science (Pérez, 2021).*
3. **Compare** with the original: have you changed the structure and vocabulary, not just a few synonyms? Do you keep the meaning?
4. **Cite** the source even though you have paraphrased.

**Bad example (*patchwriting*):** *Peer review, in spite of its limits, continues to be the principal quality-control mechanism of the scientific literature.* It changes two words and keeps the structure: it is plagiarism even if cited, because it is presented as your own wording.

**i** Example 2 · Reusing an image with a Creative Commons licence

You want to use in your work a photograph from Wikimedia Commons licensed **CC BY-SA 4.0**. Full attribution follows the **TASL** rule: *Title, Author, Source, License*:

> "UC3M Library" by Ana López, Wikimedia Commons, CC BY-SA 4.0 (link to the image and to the licence).

If you modify it (for example, crop it), say so ("cropped from the original"). With **SA** (share alike), if you publish a derivative work, it must carry the same licence. With **NC** (non-commercial) you cannot use it in a commercial context; with **ND** (no derivatives) you cannot modify it.

**i** Example 3 · Reusing your own work

You want to use in your master's thesis three pages of a paper you submitted for another course. Presenting it as new without saying so is **self-plagiarism**. The right approach: check the course rules, cite your earlier work like any other source and rework the text, or ask the teaching staff for permission if you are extending previous work.

## Graded exercises

**Basic.** Is it plagiarism to copy a sentence in quotation marks, with its reference and page?

Explanations

No: it is a correct **quotation**. It should be used sparingly and for sentences whose exact wording matters.

**Intermediate.** Paraphrase this invented text and cite it: "Reference managers reduce formatting errors and save time, but they require checking the imported metadata" (Ruiz, 2022, p. 3).

Explanations

Example: *Tools such as Zotero make it easier to format a bibliography and avoid mistakes, as long as the data they import automatically are checked (Ruiz, 2022).* Other wordings are valid if they change structure and vocabulary and keep the idea.

**Advanced.** Take a paragraph from your own work, run it through the chapter's similarity lab with a source you used, and rewrite the parts with long overlaps. Explain what changes you made.

Explanations

Overlaps of a few words (technical terms, names) are not a problem; long sequences with identical structure are. The rewriting should start from understanding, not from swapping synonyms.

## Activities

Activity: What type of plagiarism is it?

**Categorías:** Verbatim copying · Patchwriting · Uncited paraphrase · Self-plagiarism · Work done by others · Not plagiarism

- You copy two paragraphs from Wikipedia without quotation marks or citation
- You replace some words in a paragraph with synonyms and keep its structure
- You explain an article's idea in your own words, but do not cite it
- You submit for one course the work you already submitted for another
- You submit an essay generated by an AI assistant as if it were your own
- You reproduce a short sentence in quotation marks with its reference and comment on it
- You write that water boils at 100 °C at sea level without citing

Solution

- You copy two paragraphs from Wikipedia without quotation marks or citation → **Verbatim copying**
- You replace some words in a paragraph with synonyms and keep its structure → **Patchwriting**
- You explain an article's idea in your own words, but do not cite it → **Uncited paraphrase**. The idea still belongs to someone else.
- You submit for one course the work you already submitted for another → **Self-plagiarism**
- You submit an essay generated by an AI assistant as if it were your own → **Work done by others**
- You reproduce a short sentence in quotation marks with its reference and comment on it → **Not plagiarism**. It is a correct quotation, covered by the right of quotation.
- You write that water boils at 100 °C at sea level without citing → **Not plagiarism**. It is common knowledge.

ℹ Scenario: Which Creative Commons licence do you choose?

You are going to publish your own notes openly. Answer and we will tell you which licence fits (all four answers are valid: it is your decision).

**1. Do you allow commercial use?**

- Yes, commercial uses too ⟶ step 2
- No, only non-commercial uses ⟶ step 3

**2. Do you allow derivative works?**

With commercial uses allowed.

- Yes, and whoever adapts it can choose the licence ⟶ step 4
- Yes, but adaptations must carry the same licence ⟶ step 5
- No, only unmodified copies ⟶ step 6

**3. Do you allow derivative works?**

Non-commercial uses only.

- Yes, and whoever adapts it can choose the licence ⟶ step 7
- Yes, but with the same licence ⟶ step 8
- No, only unmodified copies ⟶ step 9

**4. CC BY**

*(good ending)*

The most open: it only requires attribution. It is the one used by this book and by many open access journals.

**5. CC BY-SA**

*(good ending)*

Attribution and share alike: adaptations must be just as open (like Wikipedia).

**6. CC BY-ND**

*(good ending)*

Attribution and no derivatives: it can be redistributed, commercially too, but without changes.

**7. CC BY-NC**

*(good ending)*

Attribution and non-commercial.

**8. CC BY-NC-SA**

*(good ending)*

Attribution, non-commercial and share alike.

**9. CC BY-NC-ND**

*(good ending)*

The most restrictive: only copying and distributing without changes, for non-commercial purposes and with attribution.

⚠ Common mistakes

- **"If I change the words, it is mine."** Ideas, data and structure must be cited too.
- **"If it is on the internet, it is free."** Being accessible does not mean being in the public domain or having a licence that allows reuse.
- **Citing only at the end.** The entry in the reference list is not enough: each borrowed idea needs its citation where you use it in the text.
- **Trusting the percentage** of the similarity tool, in either direction.
- **Forgetting images and figures**: they also have an author and a licence.

## Assessed challenge

! Unit 3 challenge · Writing with sources

Choose an academic or popular science article you are interested in:

1. Write a **summary** of about 150 words, a **paraphrase** of one paragraph and a short **quotation**, each with its reference.
2. Deliberately write a **patchwriting** version of the same paragraph and explain why it would be plagiarism.
3. Identify the article's **licence** and explain what uses it allows.
4. Reflect on which situations in your school or university experience might have been plagiarism without you knowing.

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student works with a different text and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. A well-written paraphrase does not need a citation because it uses your own words. *[ T / F ]*
2. Copying while changing some words but keeping the structure of the original is called _____
3. An author's moral rights cannot be waived. *[ T / F ]*
4. In a Creative Commons licence, the condition that prevents commercial use is *[ BY / SA / NC / ND ]*
5. Under the Spanish Intellectual Property Act, economic rights generally last for the author's life plus *[ 20 years / 50 years / 70 years ]*
6. Submitting work generated by an AI as your own is a form of plagiarism. *[ T / F ]*
7. Why is similarity-checking software not enough to decide whether there is plagiarism?

*(a) Because it only works with texts in English (b) Because it detects copying well, patchwriting less well and paraphrase or translation hardly at all, and it does not know whether there is a citation (c) Because it always gives false positives*

8. Reusing your own already-assessed work without saying so is called _____
9. The Creative Commons condition present in all licences is *[ BY / SA / NC / ND ]*
10. Common-knowledge facts, widely known and easy to check, do not need a citation. *[ T / F ]*

Explanations

- The idea still belongs to someone else: you must cite it even if the words are yours.
- Changing words without changing the structure.
- Recognition of authorship cannot be transferred.
- NC, non-commercial.
- 70 years after the author's death, as a general rule.
- It is presenting as your own work that is not.
- It measures text overlaps, not whether the source is used correctly.
- It is also a form of plagiarism.
- Attribution.
- When in doubt, it is better to cite.

## Further resources

| Resource | What it is for |
|---|---|
| Spanish Intellectual Property Act (BOE) | Consolidated text (in Spanish) |
| Creative Commons – Licenses | Choosing and understanding a licence |
| *The information literacy user's guide*, chapter on ethics | Open textbook |

# Units 4 and 5. References

**Objectives**

- Create and manage references in academic work.
- Apply a citation style and order the references.
- Know similarity-checking tools and their limits.

**Unit glossary**

Citation · Reference · Citation style · Author-date system · Numeric system · DOI · Reference manager · Similarity-checking tool

**Downloads**

Chapter R script Challenge submission template (.qmd) Style theme (.scss)

## Required reading

Gilmour, R.; Cobus-Kuo, L. (2011). Reference management software: A comparative analysis of four products. *Issues in Science and Technology Librarianship*, 66.

### Further reading

- Foltýnek & others (2020) — What similarity-checking tools detect and what they miss.
- American Psychological Association () — Official guide to APA 7 style.

## 1. Why cite

Citing serves to **acknowledge** other people's work, **show** what your statements are based on, **allow** readers to find the sources and **distinguish** your ideas from those of others. Academic work has two related elements:

- The **in-text citation**: a short marker where you use the source, such as (Bates, 1989) or [3].
- The **reference**: the full description of the source in the final list.

## 2. Citation styles

A **style** sets which details are included, in what order and with what punctuation. There are two main systems:

| System | In-text citation | Order of the list | Styles |
|---|---|---|---|
| **Author-date** | (Metzger & Flanagin, 2013) | Alphabetical by surname | APA 7, Harvard, Chicago author-date |
| **Numeric** | [1] or superscript | In order of appearance | Vancouver, IEEE, numeric ISO 690 |

**APA 7th edition** is widely used in this degree. What matters is to choose a style, apply it **consistently** and follow the one required by each lecturer or journal.

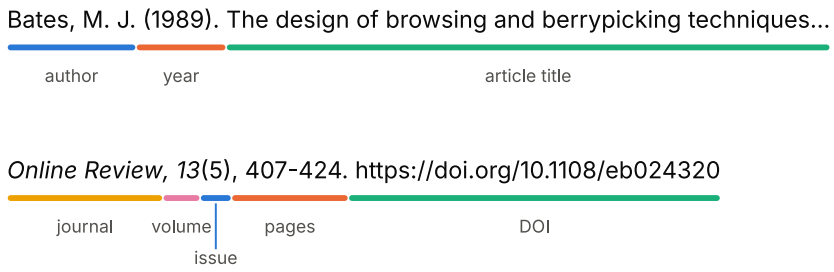


Figure 9: Anatomy of a journal article reference in APA 7 style: authors, year, article title, journal in italics, volume in italics, issue, pages and DOI as a link.

## 3. References for the most common document types (APA 7)

| Type | Pattern |
|---|---|
| Journal article | Surname, I. (Year). Title of the article. *Journal*, *volume*(issue), pages. https://doi.org/… |
| Book | Surname, I. (Year). *Title of the book* (ed.). Publisher. |
| Book chapter | Surname, I. (Year). Title of the chapter. In I. Surname (Ed.), *Title of the book* (pp. x-y). Publisher. |
| Web page | Author or organisation. (Year, Month day). *Title*. Website. URL |
| Law | Title of the law. *Official gazette*, number, date. URL |

The **DOI** (*Digital Object Identifier*) is a permanent identifier for a digital document. If there is one, include it as a link `https://doi.org/…`: even if the journal's website changes, the DOI will still lead to the document.

## 4. Reference managers

A **reference manager** stores documents with their details, organises them, inserts citations into the word processor and generates the bibliography in any style. Gilmour & Cobus-Kuo (2011) compared several managers and found that **they all make mistakes** when generating references (capitalisation, missing data, name formats), so **you must always check them**. The most widely used are **Zotero** (free and open source), Mendeley and EndNote.

A recommended workflow:

1. Capture the reference from the database or the browser (with the manager's connector).
2. Check and complete the details.
3. Organise them in collections and tags.
4. Cite from the word processor with the manager's plug-in.
5. Review the final bibliography.

## 5. Similarity-checking tools

Universities use software that compares assignments with large text databases and flags overlaps. Foltýnek & others (2020) evaluated many of these tools and concluded that they detect verbatim copying well but struggle with modifications (paraphrase, synonyms, translations), and that none is perfect. The similarity report is an **indication**, not a verdict: a well-made quotation matches its source and is not plagiarism, and an uncited paraphrase may not match and still be plagiarism.

## R practice: building and ordering references

`paste0()` joins text. With it you can see how a manager builds a reference from its data:

```r
authors <- "Bates, M. J."
year <- 1989
title <- "The design of browsing and berrypicking techniques for the online search
interface"
journal <- "Online Review"; volume <- 13; issue <- 5; pages <- "407-424"
doi <- "10.1108/eb024320"
paste0(authors, " (", year, "). ", title, ". ", journal, ", ", volume, "(", issue, "),
",
       pages, ". https://doi.org/", doi)
```

**Exercise 4.1**

In APA style, the list is ordered **alphabetically** by the first author's surname. Sort these references with `sort()` and write which one comes first.

Exercise: fill in the blanks _____

```r
sorted_refs <- _____(refs)
sorted_refs[1]
```

Hint

The function that sorts a vector is `sort()`.

Solution

```r
sorted_refs <- sort(refs)
sorted_refs[1]
```

Bates (1989): "Bat" comes before "Baw". Managers sort like this, letter by letter.

## Lab: from DOI to reference

Interactive lab

Type a DOI and the lab asks Crossref for the reference in **APA 7**, just as a reference manager does. Then **check it**: look for extra capital letters, missing italics or incomplete data. This is exactly what Gilmour & Cobus-Kuo (2011) found.

Also try: `10.1108/eb024320`, `10.1038/s41598-023-41032-5` or the DOI of the Spanish edition of this book, `10.5281/zenodo.24515144` (registered with DataCite, not Crossref).

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Worked examples

**Example 1 · An APA 7 reference for an article with a DOI, step by step**

Article details: authors G. Wilson, J. Bryan, K. Cranston, J. Kitzes, L. Nederbragt and T. K. Teal; year 2017; title "Good enough practices in scientific computing"; journal *PLOS Computational Biology*, volume 13, issue 6, article e1005510; DOI 10.1371/journal.pcbi.1005510.

1. **Authors:** surname and initials, separated by commas, with "&" before the last one (APA 7 includes up to 20 authors).
2. **Year** in parentheses.
3. **Article title** in sentence case (capital only for the first word and proper nouns), not in italics.
4. **Journal and volume in italics**, issue in parentheses, pages or article number.
5. **DOI** as a link.

> Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology, 13*(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

**In-text citation:** with three or more authors, only the first one and *et al.*: (Wilson et al., 2017) or Wilson et al. (2017).

**Example 2 · Other document types (APA 7)**

| Type | Structure |
|---|---|
| **Book** | Author, A. (Year). *Title of the book* (ed.). Publisher. |
| **Chapter** | Author, A. (Year). Title of the chapter. In E. Editor (Ed.), *Title of the book* (pp. x-y). Publisher. |
| **Web page** | Author or organisation. (Year, Month day). *Title*. Site. URL |
| **Dataset** | Author, A. (Year). *Title of the dataset* (Version) [Data set]. Repository. https://doi.org/… |
| **Software** | Author, A. (Year). *Name* (Version) [Computer software]. Repository. URL or DOI |

If a document has a DOI, always use it instead of the URL.

**Example 3 · From search to bibliography with Zotero**

1. Install Zotero and its browser connector.
2. From the article or database page, click the connector: Zotero saves the metadata (and the PDF if available). For a DOI, use the magic wand button ("Add item by identifier").
3. **Check the metadata**: title capitalisation, author names, document type. Import errors are common.
4. In the word processor, insert the citations with the Zotero plug-in and generate the bibliography in APA 7 style.
5. Check the final result against the style rules: the manager does not guarantee correct references if the source data are wrong.

## Graded exercises

**Basic.** How do you cite in the text a 2023 work by two authors, García and López?

Explanations

(García & López, 2023) in parentheses, or García and López (2023) in the narrative. With two authors, both are always cited.

**Intermediate.** Write the APA 7 reference for this dataset: author Marta Ruiz, 2024, "University reading habits survey", version 1.2, deposited in Zenodo with DOI 10.5281/zenodo.0000000 (fictitious DOI).

Explanations

Ruiz, M. (2024). *University reading habits survey* (Version 1.2) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.0000000

**Advanced.** Import into Zotero ten references on your topic from three different sources, correct their metadata and generate the bibliography in APA 7 and in another style from your field. Note the errors you found.

Explanations

Typical errors: titles in capitals, corporate authors split as people, chapters imported as books, missing DOI or pages. Noting them helps you always check what a manager imports.

## Activities

Activity: The parts of a journal article reference in APA 7

- Article title
- Authors
- DOI as a link
- Journal name, in italics
- Pages or article number
- Volume, in italics, and issue in parentheses
- Year in parentheses

Solution

1. Authors
2. Year in parentheses
3. Article title
4. Journal name, in italics
5. Volume, in italics, and issue in parentheses
6. Pages or article number
7. DOI as a link

**Activity: Author-date or numeric?**

**Categorías:** Author-date · Numeric

- (Bates, 1989)
- [3]
- List ordered alphabetically by surname
- List ordered by order of appearance
- APA 7 style
- Vancouver style
- IEEE style
- Harvard style

**Solution**

- (Bates, 1989) ⟶ **Author-date**
- [3] ⟶ **Numeric**
- List ordered alphabetically by surname ⟶ **Author-date**
- List ordered by order of appearance ⟶ **Numeric**
- APA 7 style ⟶ **Author-date**
- Vancouver style ⟶ **Numeric**
- IEEE style ⟶ **Numeric**
- Harvard style ⟶ **Author-date**

**Activity: What type of document is it?**

**Categorías:** Journal article · Book · Book chapter · Web page · Legislation

- Cruz Gil, C. (2015). *Modelos de búsqueda y recuperación de la información*. Trea.
- Bates, M. J. (1989). The design of browsing… *Online Review*, *13*(5), 407-424.
- Real Decreto Legislativo 1/1996… *Boletín Oficial del Estado*, 97.
- Caulfield, M. (2019, June 19). *SIFT (The four moves)*. Hapgood.
- Surname, I. (2020). Title. In I. Editor (Ed.), *Title of the book* (pp. 10-25). Publisher.

Solution

- Cruz Gil, C. (2015). *Modelos de búsqueda y recuperación de la información*. Trea. → **Book**
- Bates, M. J. (1989). The design of browsing… *Online Review*, *13*(5), 407-424. → **Journal article**
- Real Decreto Legislativo 1/1996… *Boletín Oficial del Estado*, 97. → **Legislation**
- Caulfield, M. (2019, June 19). *SIFT (The four moves)*. Hapgood. → **Web page**
- Surname, I. (2020). Title. In I. Editor (Ed.), *Title of the book* (pp. 10-25). Publisher. → **Book chapter**

Common mistakes

- **Mixing styles** in the same piece of work (some references in APA, others copied from the journal's website).
- **Not checking what the manager generates**: titles in capitals, abbreviated journal names or missing authors.
- **Citing the database instead of the document** ("Source: Dialnet").
- **Writing "DOI: 10…."** instead of the full link `https://doi.org/10.…`.
- **In-text citations without a reference**, or references in the list that are not cited in the text.

## Assessed challenge

Units 4 and 5 challenge · A bibliography with a reference manager

With the documents from your unit 2 challenge (or another ten on a topic of your choice):

1. Capture them in **Zotero** (or another manager) and organise them in a collection.
2. Generate the bibliography in **APA 7** and check each reference: document **all the errors** you correct.
3. Include at least one article with a DOI, a book, a chapter, a web page and a law.
4. Write a paragraph with three **in-text citations** and explain the difference between the author-date and numeric systems with your own example.

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student works with different documents and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. In the author-date system, the reference list is ordered *[ by order of appearance / alphabetically by the first author's surname / by year ]*
2. Vancouver and IEEE are styles of the *[ author-date / numeric ]* system
3. The permanent identifier of a digital document included as a link in the reference is the _____
4. The references generated by a reference manager are always correct. *[ T / F ]*
5. In APA 7, the journal title is written in italics. *[ T / F ]*
6. A similarity report of 0% guarantees that the work contains no plagiarism. *[ T / F ]*
7. What is the difference between a citation and a reference?

*(a) They are the same thing with different names (b) The citation is the short marker in the text; the reference is the full description in the final list (c) The citation goes at the end and the reference in the text*

8. In APA 7, a work by three or more authors is cited in the text with "et al." from the first citation. *[ T / F ]*

9. Which reference manager is free and open source? *[ EndNote / Zotero / Web of Science ]*
10. In a numeric system, the reference list is ordered by *[ surname / year / order of appearance in the text ]*

Explanations

- By the first author's surname.
- They number citations in order of appearance.
- *Digital Object Identifier*, permanent even if the website changes.
- Gilmour and Cobus-Kuo found errors in all of them: they must be checked.
- Journal and volume go in italics.
- An uncited paraphrase may not match and still be plagiarism.
- The citation points to the full reference.
- From the first time, unlike APA 6.
- Zotero is free software.
- The number corresponds to the order of citation.

## Further resources

| Resource | What it is for |
|---|---|
| APA Style | Official guide to APA 7 style |
| Zotero | Free reference manager |
| Crossref | DOI registration and metadata search |
| Appendix: quick citation guide | APA 7 examples from this book |

# Units 6 and 7. Generative artificial intelligence: foundations and applications

**Objectives**

- Explain what generative AI is and how language models produce text.
- Know its innovative applications, its technical challenges and open lines of research.
- Recognise its responsible uses for study and its limits, especially hallucinations.
- Verify the information and references that an AI produces.

**Unit glossary**

Generative artificial intelligence · Language model · Prompt · Hallucination · Training data · Retrieval-augmented generation

**Downloads**

Chapter R script Challenge submission template (.qmd) Style theme (.scss)

## Required reading

Walters, W. H.; Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, 13, 14045.

**Further reading**

- Kasneci & others (2023)
- Bender et al. (2021)

**Note**

The generative AI block covers units 6 to 9 of the syllabus. This chapter covers foundations, applications and technical challenges (units 6 and 7); the next one covers ethics, authorship and regulation (units 8 and 9).

## 1. What generative AI is

**Generative artificial intelligence** produces new content (text, images, audio, code) from patterns learned from large amounts of data. Conversational assistants are based on **language models**: systems trained on huge collections of texts to predict the most likely word (in fact, word fragment) that follows a given text. By repeating that prediction again and again, the model writes fluent answers.

Three important consequences follow:

- The model produces what is **plausible**, not necessarily what is **true**. Bender et al. (2021) critically called them “stochastic parrots”: they combine linguistic forms without their own access to meaning or facts.
- What it knows depends on its **training data**: it may not know recent facts and may reproduce the biases in those data.
- How you ask matters: a clear **prompt**, with context, goal and format, gives better results.

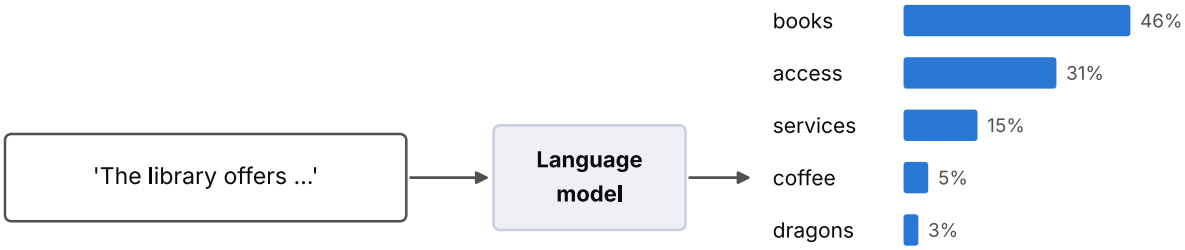


Figure 10: How a language model generates text: it estimates the probability of each possible continuation and chooses one; then it repeats. Nothing in the process checks that the result is true.

Some systems combine the model with a search engine or a document collection and answer from what they retrieve (**retrieval-augmented generation**). This improves currency and makes it possible to link sources, but it does not eliminate errors: you must check that the source says what the system claims.

## 2. Innovative applications and technical challenges

Generative AI is already used to draft and translate texts, summarise documents, write and review code, create images, audio and video, support customer service with conversational assistants and help search for and organise information. In information and documentation, applications include automatic description of collections, transcription of historical documents and search engines that answer in natural language citing their sources.

Its **technical challenges** define the open lines of research:

| Challenge | What it involves |
|---|---|
| **Truthfulness** | Reducing hallucinations and grounding answers in verifiable sources |
| **Currency** | Incorporating information after training without retraining the whole model |
| **Explainability** | Understanding why the model produces a particular answer |
| **Bias** | Detecting and mitigating biases inherited from the data |
| **Efficiency** | Reducing the energy consumption and cost of training and using the models |
| **Evaluation** | Reliably measuring what they can do and where they fail |
| **Detection** | Distinguishing generated content from human content, which cannot be done reliably today |

## 3. Hallucinations

A **hallucination** is an answer that looks correct but is false or invented. In academic work, the most dangerous one is the **fabricated reference**: an article with plausible authors, title, journal and DOI that does not exist. Walters & Wilder (2023) asked ChatGPT for short literature reviews and checked every reference: a large

share of GPT-3.5's citations were fabricated, and with GPT-4 a smaller but noticeable proportion remained; in addition, many real references contained errors.

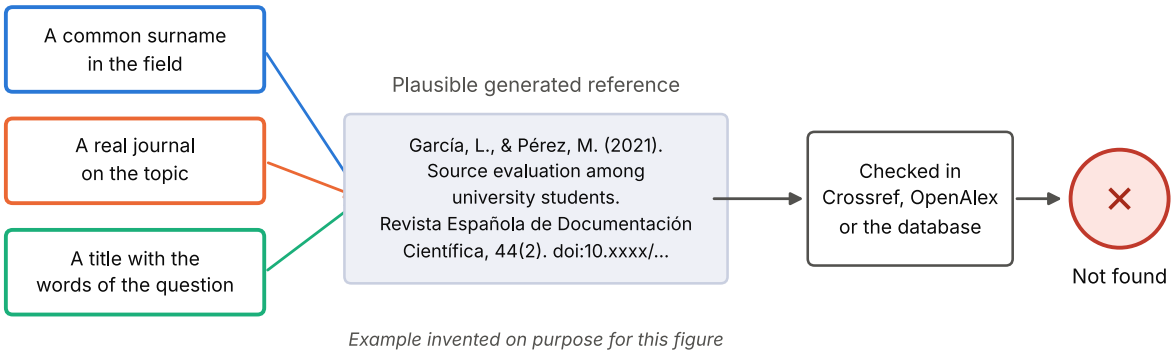


Figure 11: How a fabricated reference arises. The model combines plausible pieces (a real author, a real journal, a title on the topic) into a reference that does not exist; only checking in a database reveals it.

## 4. Responsible uses for study

Kasneci & others (2023) review opportunities (tailored explanations, writing support, personalised practice) and risks (dependence, errors, bias, plagiarism). A practical criterion is to distinguish uses by what you delegate:

| Use | Risk | Good practice |
|---|---|---|
| Asking it to explain a concept | Medium | Compare with a textbook or a reliable source |
| Generating ideas or an initial outline | Low-medium | Use it as a starting point, not as a result |
| Checking the spelling and clarity of your text | Low | Review each change; the text is still yours |
| Searching for literature | **High** | Verify **every** reference in a database |
| Summarising a document | Medium | Read the original; the summary may omit or invent |
| Writing the assignment for you | **Unacceptable** | It is presenting someone else's work as your own (Unit 3) |

Before using AI in a course, check what the teaching staff allow and declare its use when required (you will see this in the next chapter).

## 5. How to verify what an AI produces

1. **References**: look up each one by its DOI (at `doi.org` or Crossref) or by its title in Google Scholar, OpenAlex or the database for your field. Check authors, year and journal.
2. **Data and facts**: apply lateral reading and SIFT (Unit 1).
3. **Attributed quotations**: find the sentence in the original source.
4. **Content**: check that the real source says what the AI claims it says.

## R practice: checking references against a database

The `%in%` operator checks which elements of one vector are in another. Imagine an AI has given you five DOIs and you have a list of the DOIs that exist in a database:

```
ai_dois <- c("10.1108/eb024320", "10.1016/j.ipm.2019.102041", "10.1038/
s41598-023-41032-5",
             "10.1177/0165551508095781", "10.1093/jis/abc123")
database <- c("10.1108/eb024320", "10.1038/s41598-023-41032-5",
"10.1177/0165551508095781",
              "10.1016/j.pragma.2013.07.012", "10.1145/1121949.1121979")
ai_dois %in% database
```

### Exercise 6.1

Calculate the **proportion of fabricated references** (those not in the database). Remember that `!` reverses a logical value and that `mean()` of a logical vector gives the proportion of `TRUE`.

**i** Exercise: fill in the blanks _____

```
mean(!(ai_dois %in% _______))
```

**i** Hint

Compare with the vector `database`.

Solution

```
mean(!(ai_dois %in% database))
```

0.4: two of the five references do not appear. In practice, a DOI missing from a list does not prove it does not exist: you have to look it up at `doi.org`. But a DOI with a strange format or one that does not resolve is a clear sign of a fabricated reference.

## Lab: a miniature language model

**i** Interactive lab

This toy model has "read" only a dozen sentences about libraries and information. For each word it counts which words follow it and how often (a **bigram model**). Choose a starting word and the **temperature**: with low temperature it almost always picks the most likely word; with high temperature it takes more risks. Generate several sentences: you will see fluent and **plausible** texts, some of them false. Large models are vastly more complex, but they share this idea: predicting what comes next, not checking the truth.

**Try:** generate sentences from "the" at temperature 2. Do any of them claim something the corpus does not say? That is, on a small scale, a hallucination.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Lab: verifying references

ℹ Interactive lab

Paste the DOIs an AI has given you (one per line; the full link also works). The lab checks at **doi.org** whether each DOI exists and, if it is registered with Crossref, retrieves the title and year so you can check that they match what the AI told you.

A DOI that exists is not enough: check that **the title, authors and year** match the reference the AI gave you. Walters & Wilder (2023) found, as well as fabricated references, many real references with errors in their details.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Lab: review your prompt

ℹ Interactive lab

A good prompt usually gives **context**, a specific **task**, the **format** of the answer, **limits** and a request for **verification**. Write yours: the lab looks for those pieces by keywords and suggests what to add. It is a rough guide, not a mark.

Try to improve the example prompt until it has all five pieces and compare it with the one in example 2.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Worked examples

ℹ Example 1 · Auditing a bibliography generated by an AI

You ask an assistant for five references on "reproducibility in psychology". Procedure for each reference:

1. **Does it have a DOI?** Open it at `https://doi.org/` followed by the DOI. If it does not resolve, be suspicious.
2. **If it resolves, compare** title, authors, journal and year with what the AI gave: sometimes the DOI exists but belongs to another article.
3. **If it has no DOI**, search for the exact title in quotation marks in Google Scholar, Crossref or OpenAlex.
4. **Classify** each reference: correct, with errors (year, authors, pages) or **fabricated**.
5. **Do not use** any reference you have not located and read at least in its abstract.

**Common result:** some references are real, others mix the authors of one work with the title of another, and one does not exist. This is what is known as a **hallucination**: the model generates plausible text, it does not query a database (unless it uses a search tool, and even then it is worth checking).

Example 2 · Improving a prompt

**Vague prompt:** “Tell me about open science”.

**Improved prompt:** > Act as a tutor for an undergraduate student of Information Science. Explain in 300 words what open science is, with three examples of practices (open access, open data, preregistration). Use clear language. Do not include references: I will search for them myself in databases. At the end, suggest three questions to check whether I have understood.

Elements that improve the result: **role**, **audience**, **specific task**, **length**, **format** and **limits** (here, avoiding references that could be fabricated).

Example 3 · A figure that sounds right

An AI states: “70% of psychology studies cannot be replicated, according to the 2015 project”. Verification: the Open Science Collaboration (2015) project replicated 100 studies and obtained significant results in the same direction in 36% of the replications, compared with 97% of the originals. The AI’s figure is a simplification that mixes criteria: you need to go to the source and cite the exact figure with its definition.

## Graded exercises

**Basic.** Why can a language model fabricate a reference that looks correct?

Explanations

Because it generates the most likely sequence of words from patterns in its training: it produces something shaped like a reference (plausible authors, a real journal, a year) without checking that it exists.

**Intermediate.** Rewrite this prompt so that it is useful and safe: “Write the literature review for my bachelor’s thesis”.

Explanations

Example: “My bachelor’s thesis is about [topic]. Propose a five-section structure for the literature review, with the key concepts of each section and the search terms I should use in Dialnet and Scopus. Do not write the text and do not invent references”. This way the AI helps you plan and does not replace your work.

**Advanced.** Ask two different AI assistants for ten references on a topic you are interested in, audit them all with the procedure in example 1 and compare the error rates.

Explanations

Present a table with each reference, its status (correct, with errors, fabricated) and the evidence (DOI, search). State the date and the model used: results change between versions.

## Activities

**Activity: How risky is this use?**

**Categorías:** Low · Medium · High · Unacceptable

- Asking it to check the spelling of a text you wrote
- Asking it to explain a concept you do not understand
- Asking for a list of references for your assignment and copying it without checking
- Asking it to summarise an article you are not going to read
- Submitting work written by the AI as your own
- Generating ideas to choose the topic of your assignment
- Uploading your group members' personal data so that it assigns tasks

**Solution**

- Asking it to check the spelling of a text you wrote → **Low**
- Asking it to explain a concept you do not understand → **Medium**. Check the explanation against a reliable source.
- Asking for a list of references for your assignment and copying it without checking → **High**. Many may be fabricated.
- Asking it to summarise an article you are not going to read → **Medium**. The summary may omit or invent.
- Submitting work written by the AI as your own → **Unacceptable**
- Generating ideas to choose the topic of your assignment → **Low**
- Uploading your group members' personal data so that it assigns tasks → **High**. It puts other people's privacy at risk.

**Activity: Which technical challenge is it?**

**Categorías:** Truthfulness · Currency · Explainability · Bias · Efficiency · Detection

- The assistant does not know who won an election last month
- Nobody can explain why the model chose that answer
- The model associates "nurse" with women and "surgeon" with men
- Training a large model uses as much electricity as hundreds of homes in a year
- A lecturer cannot know for sure whether a text was written by an AI
- The model cites an article that does not exist

Solution

- The assistant does not know who won an election last month → **Currency**
- Nobody can explain why the model chose that answer → **Explainability**
- The model associates "nurse" with women and "surgeon" with men → **Bias**
- Training a large model uses as much electricity as hundreds of homes in a year → **Efficiency**
- A lecturer cannot know for sure whether a text was written by an AI → **Detection**
- The model cites an article that does not exist → **Truthfulness**

Common mistakes

- **Using AI as a literature search engine** without verifying each reference.
- **Asking the AI itself whether the reference exists**: it may confirm it with complete confidence even if it is false.
- **Taking fluency for accuracy**: a well-written text is not a true text.
- **Delegating what the course wants you to learn.** AI can help you learn, not learn for you.
- **Not declaring its use** when the teaching staff require it (next chapter).

## Assessed challenge

! Units 6 and 7 challenge · Auditing an AI's references

1. Ask a generative AI assistant for a list of **ten academic references** on a topic you are interested in. Save the full prompt and answer.
2. Verify **each reference**: does it exist? Are the authors, year, title, journal and DOI correct? Does the content match the topic?
3. Classify each one as **correct**, **real with errors** or **fabricated**, and calculate the proportions.
4. Repeat the search in a database (Unit 2) and compare the quality of the results and the time taken.

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student works on a different topic and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. A language model generates text by predicting the most likely fragment that follows a given text. *[ T / F ]*
2. An AI answer that looks correct but is false or invented is called a _____
3. If an AI gives a reference with a DOI, authors and journal, the reference exists. *[ T / F ]*
4. Which of these uses is riskiest? *[ Checking the spelling of your text / Asking for initial ideas / Searching for literature without verifying it ]*
5. If 3 of 12 references generated by an AI are fabricated, the proportion fabricated is *[ 0.12 / 0.25 / 0.33 / 3 ]*
6. Retrieval-augmented generation completely eliminates errors. *[ T / F ]*
7. Why can a language model fabricate a plausible reference?

*(a) Because it copies references from a database with errors (b) Because it generates plausible text by combining patterns, without checking that the reference exists (c) Because its goal is to deceive the user*

8. What a language model knows depends on its _____ data

9. Detectors exist today that distinguish AI-generated text from human text with complete reliability. *[ T / F ]*
10. To check whether a reference given by an AI exists, the most reliable approach is to *[ ask the same AI whether it exists / look it up by its DOI or title in Crossref, OpenAlex or a database / check that the format looks correct ]*

Explanations

- It repeats that prediction to write the answer.
- A plausible but false answer.
- It may be fabricated: it must be verified.
- Walters and Wilder found many fabricated references.
- 3 / 12 = 0.25.
- It improves currency, but sources must be checked.
- It produces what is plausible, not what is true; there is no intent.
- It may not know recent facts and may inherit their biases.
- Detection is one of the open technical challenges.
- The AI may confidently confirm a false reference.

## Further resources

| Resource | What it is for |
|---|---|
| Crossref | Checking DOIs and metadata |
| OpenAlex | Checking whether an article exists |
| UNESCO – Guidance for generative AI in education and research (Miao & Holmes, 2023) | International guidance |

# Units 8 and 9. Generative artificial intelligence: ethics, authorship and regulation

**Objectives**

- Analyse the ethical principles of generative AI: consent, privacy and informational self-determination.
- Understand responsibility and authorship in the use of AI.
- Declare the use of AI transparently in academic work.
- Know the European regulatory framework and the ethical principles of AI.

**Unit glossary**

Algorithmic bias · AI use statement · AI Act · AI literacy · Explainability · Personal data · Informational self-determination

**Downloads**

Challenge submission template (.qmd) Style theme (.scss)

## Required reading

Floridi, L.; Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).

### Further reading

- Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union.*
- Bender & others (2021) — Environmental costs and biases of large models.

## 1. Ethical principles of AI

Floridi & Cowls (2019) compared the main international statements on AI ethics and summarised them in **five principles**:

| Principle | What it means |
|---|---|
| **Beneficence** | Promoting the well-being of people and the planet |
| **Non-maleficence** | Avoiding harm: privacy, security, misuse |
| **Autonomy** | People keep the power to decide |
| **Justice** | Sharing benefits, avoiding discrimination |
| **Explicability** | Being able to understand how it works and who is accountable |

The first four come from bioethics; the fifth is specific to AI and makes the others possible.

## 2. Ethical risks in everyday use

- **Bias**: models reproduce the biases in their training data (gender, origin, language). An answer can look neutral and be biased underneath.
- **Consent, privacy and informational self-determination**: **informational self-determination** is each person's right to decide what is done with their personal data; in Spain, the Constitutional Court recognises it as the fundamental right to data protection. Models are often trained on data collected from the internet

without their authors' consent, and what you write into an assistant may be stored and, depending on the service, used to train models. Do not enter your own or other people's personal data, or confidential documents, and check the service's privacy settings.

- **Authorship and intellectual property**: a generated text is not yours; moreover, models are trained on protected works, an open legal debate.
- **Dependence and learning**: delegating to AI what the course wants you to learn prevents you from learning it.
- **Environmental cost**: training and using large models consumes a lot of energy and water (Bender et al., 2021).

## 3. Responsibility, authorship and transparency

In academic work, the general rule is **transparency**: if you use AI, say so, explain what for and take responsibility for the result. Scientific publishers and COPE (Committee on Publication Ethics) agree that an AI **cannot be an author**, because it cannot take responsibility for the work.

A useful statement says **which tool** (and version), **what for** it was used, **in which parts** and **how** the result was checked. For example:

> AI use statement: I used [tool, version] to check the spelling and clarity of sections 2 and 3, and to suggest synonyms in the table of search terms. I reviewed and accepted or rejected each suggestion. The searches, the selection of documents and the conclusions are my own. I did not enter any personal data.

If the AI influences the content, APA also recommends citing it in the text and in the references, giving the prompt used in an appendix or in the text.

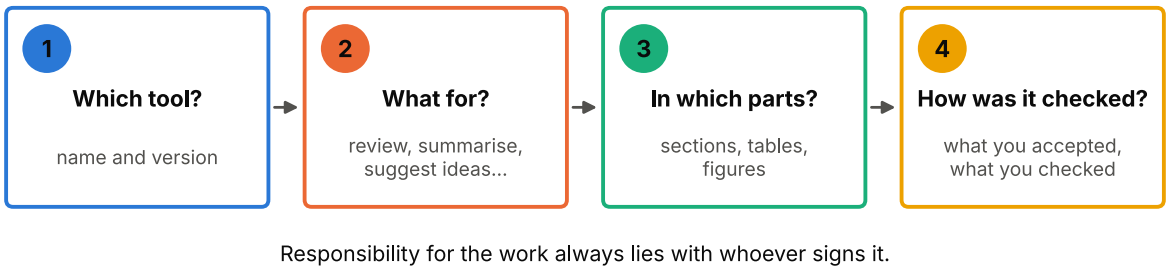


Figure 12: From use to statement: which tool, what for, in which parts and how it was checked.

## 4. The European regulatory framework

**Regulation (EU) 2024/1689** on Artificial Intelligence (*AI Act*) is the first comprehensive law on AI. It follows a **risk-based approach**:

| Risk level | Examples | Consequence |
|---|---|---|
| **Unacceptable** | Social scoring, manipulation causing harm | Prohibited |
| **High** | AI in education to assess or admit students, in employment, in essential services | Strict requirements: risk management, quality data, human oversight |
| **Transparency** | Conversational assistants, generated content (*deepfakes*) | Informing people that they are interacting with an AI or that the content is artificial |

| Risk level | Examples | Consequence |
|---|---|---|
| **Minimal** | Spam filters, video games | No specific obligations |

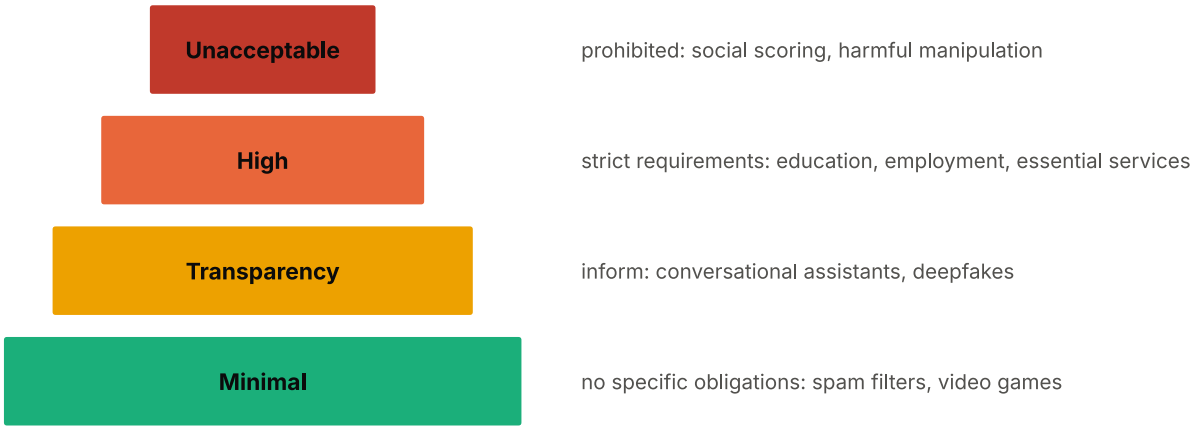


Figure 13: The four risk levels of the EU AI Act.

The regulation also covers **general-purpose AI models** (the ones behind assistants) and requires those who deploy AI systems to ensure that their staff have a sufficient level of **AI literacy**. Its obligations apply in stages from 2025. Other rules, such as the General Data Protection Regulation, continue to apply to AI.

## Worked examples

**i** Example 1 · Writing an AI use statement

In your bachelor's thesis you used an AI assistant to check spelling and to suggest an initial structure for the table of contents. A suitable statement:

> During the preparation of this work, [name of the tool, version] was used to check the spelling and grammar of the text and to obtain an initial proposal for the structure of the table of contents, which was modified by the author. All content was written, reviewed and verified by the author, who takes responsibility for the work. AI was not used to generate results, analyses or references.

Key elements: **which tool**, **for which tasks**, **what you did** and **responsibility**. Always check your university's and your course's rules.

**i** Example 2 · Personal data in a chatbot

To analyse the open-ended answers of a survey, you plan to paste them into a commercial chatbot. They contain names and comments about the participants' health. Problems: they are **personal data**, some of them in a **special category** (health), and the GDPR requires a legal basis, information to participants and guarantees about where and how the data are processed. Alternatives: **anonymise** first (remove names and identifiable information), use tools approved by the university with contractual guarantees or do the analysis locally. When in doubt, ask the data protection officer.

Example 3 · Which risk category? (EU AI Act)

Regulation (EU) 2024/1689 classifies uses of AI by their risk:

| Category | Examples | Consequence |
|---|---|---|
| **Unacceptable risk** | Social scoring of people; manipulation that causes harm | Prohibited |
| **High risk** | AI to decide access to education or to assess students; recruitment | Strict requirements: risk management, data, human oversight, transparency |
| **Transparency risk** | Chatbots; generated or manipulated content (deepfakes) | Informing people that they are interacting with an AI or that the content is artificial |
| **Minimal risk** | Spam filters, video games | No specific obligations |

A system that automatically marks exams and decides the grade would be in the **high-risk** category.

## Graded exercises

**Basic.** Who is responsible for the content of work done with the help of an AI?

Explanations

The author. The AI cannot be an author or take responsibility; whoever signs the work answers for its accuracy and originality.

**Intermediate.** Classify according to the categories of the Regulation: (a) a student support chatbot; (b) a system that decides admission to a master's programme; (c) a spell checker.

Explanations

(a) Transparency risk: it must say that it is an AI. (b) High risk: it affects access to education. (c) Minimal risk.

**Advanced.** Analyse a real or realistic case of AI use in your academic environment with the chapter's ethical principles and propose a short policy for your course or group.

Explanations

A good policy says which uses are allowed and which are not, how they are declared, what happens to personal data and how the work is verified. It must be consistent with the university's rules.

## Activities

**Activity: What risk level does the AI Act assign?**

**Categorías:** Unacceptable (prohibited) · High · Transparency · Minimal

- A system that scores citizens on their social behaviour to grant or deny them services
- A system that decides student admission to a master's programme
- A system that marks and grades university exams
- A conversational assistant on a city council website
- An AI-generated video imitating a real person (*deepfake*)
- A spam filter
- A system that screens CVs in a recruitment process
- A system that subliminally manipulates people and causes them harm

**Solution**

- A system that scores citizens on their social behaviour to grant or deny them services → **Unacceptable (prohibited)**
- A system that decides student admission to a master's programme → **High**
- A system that marks and grades university exams → **High**
- A conversational assistant on a city council website → **Transparency**. It must say that you are talking to an AI.
- An AI-generated video imitating a real person (*deepfake*) → **Transparency**. It must be labelled as artificial.
- A spam filter → **Minimal**
- A system that screens CVs in a recruitment process → **High**
- A system that subliminally manipulates people and causes them harm → **Unacceptable (prohibited)**

**Activity: Which of Floridi and Cowls' principles is at stake?**

**Categorías:** Beneficence · Non-maleficence · Autonomy · Justice · Explicability

- A grant system rejects a student and nobody can explain why
- A recruitment model penalises women's CVs
- An assistant leaks users' personal data in its answers
- An app decides which courses you take without you being able to change it
- A system helps detect early school drop-out in time to offer support

Solution

- A grant system rejects a student and nobody can explain why → **Explicability**
- A recruitment model penalises women's CVs → **Justice**
- An assistant leaks users' personal data in its answers → **Non-maleficence**
- An app decides which courses you take without you being able to change it → **Autonomy**
- A system helps detect early school drop-out in time to offer support → **Beneficence**

Scenario: Does it need to be declared?

You are finishing the final assignment for a course. The course guide asks you to declare any use of generative AI.

**1. What you have done**

You used an assistant to correct your spelling and suggest synonyms in the table of search terms. You did everything else yourself.

- I do not declare it: it was only spelling → step 2
- I declare it, stating the tool, what for and in which parts → step 3

**2. Not declared**

*(bad ending)*

The guide asks you to declare **any** use. Even if the use is minor and legitimate, hiding it breaks the course rules and may be treated as a breach of integrity.

**3. The statement**

How do you write it?

- "I used AI." → step 4
- "I used (tool and version) to check spelling and suggest synonyms in the table in section 2. I reviewed each change; the search, the selection and the conclusions are my own." → step 5

**4. Too vague**

*(could be better)*

Readers do not know which part is yours. State which tool, what for, in which parts and how you checked the result.

**5. Complete statement**

*(good ending)*

Transparent, specific and taking responsibility for the result: exactly what scientific publishers and COPE ask for.

⚠ Common mistakes

- **Thinking that "declaring" means "confessing".** Declaring a permitted use is good practice, not misconduct.
- **Believing the statement shifts responsibility to the tool.** Whoever signs answers for the whole work.
- **Entering personal or confidential data** into services that may use it to train models.
- **Assuming the AI Act only affects technology companies**: it also places obligations on those who use (deploy) high-risk systems, such as a university.

## Practice

Find UC3M's or your course's AI use policy and the AI policy of a journal in your field. Compare them: what do they allow, what do they prohibit and what do they ask you to declare?

## Assessed challenge

! Units 8 and 9 challenge · An ethical case and a use statement

1. Choose a **real case** of generative AI use with ethical problems (in education, journalism, science or culture), documented in at least two reliable sources.
2. Analyse it with Floridi and Cowls' **five principles**.
3. Place it in the **risk levels** of the EU AI Act and explain which obligations would apply.
4. Write the **AI use statement** for your own report (even if you have not used any, say so explicitly).

**Submission:** a report made with the submission template. It is assessed with the rubric included in the template. This challenge **has no solution in the book**: each student analyses a different case and the teaching staff mark it.

## Self-assessment

Answer all the questions and press **Show answers**: correct answers turn green, wrong ones red, and you will see your score. You can change your answers and check again. It does not count towards your mark.

1. The principle that Floridi and Cowls add to the four from bioethics is _____
2. An AI can be listed as co-author of a scientific article if it has written much of the text. *[ T / F ]*
3. Under the EU AI Act, an AI system used to assess students is *[ minimal / transparency / high / unacceptable ]* risk
4. Social scoring of people is *[ allowed with oversight / prohibited / subject only to transparency ]*
5. It is good practice to enter classmates' personal data into an AI assistant so it can write a group assignment. *[ T / F ]*
6. The EU AI Act follows an approach based on *[ model size / risk / economic sector ]*
7. What should an AI use statement include?

*(a) Only the name of the tool (b) Which tool, what for, in which parts and how the result was checked (c) The full conversation with the AI in every case*

8. Each person's right to decide on the use of their personal data is called informational _____
9. Informing people that content has been generated by AI is an obligation of the *[ minimal / transparency / unacceptable ]* risk level
10. If you declare your use of AI, you are no longer responsible for any errors in the work. *[ T / F ]*

Explanations

- It makes the other four principles possible.
- Authorship requires responsibility, and an AI cannot take it on.
- Education and the assessment of students are high-risk areas.
- It is a prohibited practice.
- It puts other people's privacy at risk.
- Obligations grow with the risk.
- This way readers know what is yours and what is not.
- The basis of the right to data protection.
- It affects assistants and synthetic content.
- Responsibility remains with whoever signs.

## Further resources

| Resource | What it is for |
|---|---|
| Regulation (EU) 2024/1689 | Official text of the AI Act |
| COPE – Authorship and AI tools | Position on AI and authorship |
| APA Style – How to cite ChatGPT | Citing AI-generated content |
| UNESCO – Recommendation on the Ethics of AI | International ethical framework |

# Quick guide to APA 7 citations

Examples of references in **APA style, 7th edition**, with documents that appear in this book. For cases not covered here, see the official APA Style guide and follow your lecturer's instructions.

## In-text citations

| Case | Parenthetical citation | Narrative citation |
|---|---|---|
| One author | (Bates, 1989) | Bates (1989) |
| Two authors | (Metzger & Flanagin, 2013) | Metzger & Flanagin (2013) |
| Three or more | (Sugimoto et al., 2017) | Sugimoto et al. (2017) |
| Organisation | (American Psychological Association, 2020) | American Psychological Association (2020) |
| Direct quotation | (Bates, 1989, p. 410) | Bates (1989, p. 410) |
| Several works | [Bates (1989); Marchionini, 2006) | — |

Use "&" between two authors inside parentheses and "and" in narrative citations. Quotations of fewer than 40 words go in quotation marks; longer ones go in an indented block without quotation marks.

## References

**Journal article with a DOI**

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210-220. https://doi.org/10.1016/j.pragma.2013.07.012

**Article with an article number instead of pages**

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13*, Article 14045. https://doi.org/10.1038/s41598-023-41032-5

**Book**

Cruz Gil, C. (2015). *Modelos de búsqueda y recuperación de la información*. Trea.

**Edited book, online**

Bobish, G., & Jacobson, T. (Eds.). (2014). *The information literacy user's guide: An open, online textbook*. Open SUNY Textbooks. https://milnepublishing.geneseo.edu/the-information-literacy-users-guide-an-open-online-textbook/

**Conference paper in proceedings**

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. https://doi.org/10.1145/3442188.3445922

**Law (Spain)**

Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual [Royal Legislative Decree 1/1996 approving the consolidated text of the Intellectual Property Act]. *Boletín Oficial del Estado*, 97, 22 April 1996. https://www.boe.es/buscar/act.php?id=BOE-A-1996-8930

**EU legislation**

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*, L, 12 July 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

**Web page**

Caulfield, M. (2019, June 19). *SIFT (The four moves).* Hapgood. https://hapgood.us/2019/06/19/sift-the-four-moves/

**Generative AI tool** (when its output is incorporated into the work)

Company. (Year). *Name of the tool* (version) [Large language model]. URL

Include the prompt used in the text or in an appendix, because readers cannot retrieve the conversation.

## Rules most often forgotten

- The list is ordered **alphabetically** by the first author's surname; two works by the same author, by year.
- **Hanging indent** (the first line of each reference sticks out) and double spacing in formal work.
- The **article title** is in roman type and sentence case; the **journal** and **volume** are in italics; the issue is in parentheses, in roman type.
- Write the DOI as a full link: `https://doi.org/…`, without "DOI:" in front.
- Up to **20 authors** are all listed in the reference; in the in-text citation, use "et al." for three or more from the first citation.
- Every in-text citation has its reference in the list, and every reference in the list is cited in the text.

# Checklists

Short lists to keep at hand. You can print them or copy them into your notes.

## Before believing or sharing something

- ☐ Do I know who is behind the source? (If not, find out away from the page: lateral reading.)
- ☐ What are their interests: to inform, sell, persuade?
- ☐ Do other reliable, independent sources report it, or do they all copy the same thing?
- ☐ Have I got to the original source (the study, the data, the law)?
- ☐ Does the headline say the same as the original, or does it exaggerate?
- ☐ Does the date matter? Is it still valid?

## Before you start searching

- ☐ I have written my question in one sentence.
- ☐ I have separated the concepts and have synonyms and variants for each.
- ☐ I have decided the limits: dates, languages, document types.
- ☐ I know which platforms I will use and why.
- ☐ I have somewhere to note the strategy and the results.

## Before submitting an assignment

- ☐ Every idea, figure or piece of data from others has its in-text citation.
- ☐ Every citation in the text has its reference, and vice versa.
- ☐ Quotations are in quotation marks and have a page number.
- ☐ I have used a single citation style, consistently.
- ☐ I have checked the references generated by the reference manager.
- ☐ Images that are not mine have an author and a licence.
- ☐ I have declared the use of AI, if there was any and the course requires it.

## Before using generative AI

- ☐ I know what the teaching staff allow in this activity.
- ☐ I will not enter personal data or confidential documents.
- ☐ I will verify every fact and every reference it gives me.
- ☐ I review, understand and take responsibility for the final result.
- ☐ I note the tool, the version and what I use it for, in case I need to declare it.

# Verification lab

This appendix complements unit 1 with a timeline of disinformation, a tool for evaluating sources and an activity for spotting red flags.

## Timeline: disinformation, science and regulation

## Lab: evaluate a source with the CRAAP test

ℹ Interactive lab

The **CRAAP test**, created at the Meriam Library of California State University, Chico, evaluates a source with five criteria. Score a source you are using (from 0 to 10) and look at its profile.

Criteria below 5 are marked with ⚠️ . A high score does not replace critical judgement: the test is a guide, not a formula.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Activity: red flag or not?

Read this fictitious post and classify each detail.

ℹ “DEFINITIVE STUDY! Dark chocolate cures insomnia”

Published on *TotalHealth24.info* · unsigned · date: “3 years ago”, shared 40,000 times today. The article links to a study with 12 participants funded by a chocolate brand. At the end there is a discount code to buy “therapeutic chocolate”.

ℹ Activity: Classify each detail of the post

**Categorías:** Red flag · Neutral detail

- The article is not signed
- The headline says “definitive” and uses capitals
- The study has 12 participants
- The study is funded by a chocolate brand
- It includes a discount code
- It was published 3 years ago and is circulating today
- The article links to the original study
- 40,000 people have shared it

Solution

- The article is not signed → **Red flag**. Authorship cannot be checked (IFLA: check the author).
- The headline says "definitive" and uses capitals → **Red flag**. Sensationalist language: science is rarely definitive with a single study.
- The study has 12 participants → **Red flag**. Such a small sample does not allow generalisation.
- The study is funded by a chocolate brand → **Red flag**. Possible conflict of interest.
- It includes a discount code → **Red flag**. Its purpose is to sell (CRAAP: purpose).
- It was published 3 years ago and is circulating today → **Red flag**. Check the date: later studies may have superseded it.
- The article links to the original study → **Neutral detail**. It is good practice; now you need to read the study and check what it really says.
- 40,000 people have shared it → **Neutral detail**. Popularity says nothing about whether it is true.

## Lab: misleading charts

Interactive lab

The same data can tell different stories depending on how they are drawn. These are a library's (fictitious) monthly loans. Change the chart's design decisions and see how much loans appear to grow.

In **bar** charts the axis must start at 0, because we compare lengths. In **line** charts the axis can be narrowed if this is clearly indicated, because we compare slopes.

This lab is interactive in the web version: pablosancheznunez.com/digital-skills-information

## Scenario: is this photo real?

ℹ Scenario: A widely shared photo

You receive by messaging app a photo of a flood in a library with the text "This is what your campus library looks like today". It has been forwarded many times.

**1. First reaction**

What do you do before forwarding it?

- I do a reverse image search → step 2
- I forward it with "Does anyone know if this is true?" → step 5
- I trust it: someone I trust sent it → step 6

**2. Reverse image search**

The reverse search (for example, with Google Lens, Bing or TinEye) shows that the photo was already circulating four years ago with another caption. What do you do now?

- I look for the earliest post and its original source → step 3
- I conclude it is fake and say so in the group, nothing more → step 4

**3. Back to the origin**

*(good ending)*

The earliest post is from a local newspaper in another city and shows a different library. You reply in the group with a link to the original source and a neutral sentence: "The photo is real, but it is of another library and from 2021". You have applied SIFT: **stop, investigate the source, find better coverage and trace the original context**.

**4. Hasty conclusion**

*(could be better)*

You are right that the message is misleading, but "it is fake" is not accurate: the photo is real and **taken out of context**. Providing the original source is more convincing than a statement.

**5. Forwarding with a question**

*(bad ending)*

Even if you ask, forwarding amplifies the hoax: many people will see the photo and not the question. Check before sharing.

**6. Trusting the sender**

*(bad ending)*

The person may be trustworthy and have been deceived. Trust in whoever shares something does not replace verifying the source.

## Practise with real cases

- Full Fact (UK), Maldita.es and Newtral (Spain) publish fact-checks; they are part of the International Fact-Checking Network (IFCN).
- The European Digital Media Observatory (EDMO) brings together research and fact-checks from across the EU.
- The IFLA infographic has been translated into dozens of languages (37 according to IFLA).

## Watch and keep learning

These are links, not embedded videos: the video opens on the publisher's website, so this page does not load third-party trackers.

- **Check Yourself with Lateral Reading (Crash Course Navigating Digital Information #3)** — CrashCourse, with the Stanford History Education Group and MediaWise. About 14 minutes on lateral reading; part of a 10-episode series (2019).

# Guide for teaching staff

This guide explains how the book is designed for teaching the course and how to maintain it.

## Correspondence with the course description

The six chapters follow, in the same order, the four blocks and nine units of the official syllabus of **Digital Skills for Information Use** (*Competencias digitales para el uso de la información*, 19408).

| Chapter | Units | Block |
|---|---|---|
| Information sources | 1 | Introduction. Information sources |
| Information retrieval | 2 | Retrieval and ethical use |
| Ethical use of information and plagiarism | 3 | Retrieval and ethical use |
| References | 4 and 5 | References |
| Generative AI: foundations and applications | 6 and 7 | Generative artificial intelligence |
| Generative AI: ethics, authorship and regulation | 8 and 9 | Generative artificial intelligence |

| Assessment activity (course description) | Weight | Support in the book |
|---|---|---|
| Practical assignments and presentations | 70% | **Assessed challenges** (one per chapter) |
| Final test (minimum 4) | 30% | Contents and required readings; the self-assessments and the Moodle XML question bank serve as practice |
| Resit (multiple-choice test 50% and practical assignment 50%) | — | Question bank and any of the challenges |

## Assessed challenges

Each chapter ends with a challenge with no hint, solution or automatic marking. Each student chooses their own topic, claim or case, which makes copying harder. Challenges 2 and 4 are linked: the documents found in the search strategy are then managed in Zotero. They are submitted with the `entrega_reto.qmd` template, which includes a common rubric: searching and selecting sources (25%), critical evaluation (25%), ethical use and citation (20%), argumentation (20%) and presentation (10%). The template produces an HTML file with the images embedded, so the submission is a single file; it can also be submitted in the format the teaching staff indicate. No knowledge of R is needed to use it.

Reference solutions and marking criteria for the challenges are not in the public repository.

## Self-assessment in Aula Global

Each chapter's self-assessment questions are exported to **Moodle XML** (*Question bank → Import → Moodle XML format*): one file per chapter and one with the whole bank. Each question keeps its explanation as general feedback, and the bank can be used to prepare the final test and the resit.

**Download the question bank:** whole book (Moodle XML) · by chapter: 00 · 01 · 02 · 03 · 04 · 05 · 06

## Interactive activities and labs

As well as the self-assessment, each chapter has **activities with instant feedback** (ordering, classifying and decision scenarios) and several **labs**: searching OpenAlex with Boolean operators, comparing texts like a similarity checker, turning a DOI into an APA reference, verifying the DOIs an AI gives you and playing with a miniature language model. Labs with live data query open APIs (OpenAlex, Crossref and doi.org) directly from the student's browser; nothing is sent to the book's server. In the PDF and EPUB, the activities appear in printed form, with their solution.

The activities are written in Markdown with the `interactivos` extension (see `_extensions/interactivos/interactivos.lua`), so they can be added or changed without programming.

## The role of R

The book uses as little R as possible: four chapters have a short fill-in-the-blanks exercise that shows how a tool works inside (Boolean operators, ordering references, detecting overlaps, verifying DOIs). The chapter Using R in this book is enough preparation. No challenge requires programming.

## Book features

- **Class slides**: each chapter has slides with the objectives, figures, main tables, the reading, discussion questions and the challenge. They are in diapositivas/; press F for full screen and S to see the notes.
- **Student progress**: each chapter shows the exercises solved and the questions answered correctly; the welcome page summarises them all. It is stored only in the student's browser (`localStorage`): teaching staff cannot see it.
- **Offline use**: a *service worker* (`sw.js`) stores the pages and the R engine after the first visit.
- **Light and dark mode**, with accessibility audits (WCAG 2.1 AA) in both.
- **PDF and EPUB** generated with `python3 herramientas/pdf.py`.

## Maintenance

From the root of the repository:

```
python3 herramientas/preparar.py        # sw.js and continuous integration workflow
python3 herramientas/extraer_r.py       # .R scripts for each chapter
python3 herramientas/moodle.py          # question bank
python3 herramientas/figuras.py         # figures
python3 herramientas/diapositivas.py    # class slides (after quarto render)
python3 herramientas/pdf.py             # PDF and EPUB
quarto render
```

The `.github/workflows/comprobar.yml` workflow renders the book and checks all the exercises, questions and accessibility in light and dark mode in a real browser, and once a week it checks the external links. When you publish a new version, update `version` in `CITATION.cff` and run `preparar.py` again so that browsers refresh their copy.

The data in the exercises are **invented for teaching purposes**; there are no data on real people.

This book is the English edition of *Competencias digitales para el uso de la información.*

# Glossary

Terms in the book, ordered by chapter. Each entry links to the chapter where it is explained.

## Information sources

**Information overload**
A situation in which you receive more information than you can process, with effects such as anxiety or feeling less well informed. → Unit 1

**Disinformation**
False information created or spread with the intention to deceive. → Unit 1

**Misinformation**
False information shared without the intention to deceive. → Unit 1

**Credibility**
The degree to which a source or a message is perceived as trustworthy. Online it is often judged with mental shortcuts (heuristics) such as reputation or endorsement by others. → Unit 1

**Lateral reading**
A verification strategy that consists of leaving the page and consulting other sources about it and about what it claims, instead of analysing it only from the inside. → Unit 1

**SIFT method**
Mike Caulfield's method for verifying information in four steps: stop, investigate the source, find better coverage and trace claims back to the original. → Unit 1

**Primary source**
A source containing original information: a research article, a dataset, a law or an interview. Secondary sources analyse or summarise it; tertiary sources help you find it. → Unit 1

**Peer review**
Evaluation of a piece of work by specialists in the field before it is published. → Unit 1

**Altmetrics**
Measurement of mentions of scientific publications on social media, in the news and in other online sources. It measures attention, not quality. → Unit 1

## Information retrieval

**Information need**
What a person wants to know and why; it is the starting point of every search. → Unit 2

**Boolean operator**
A word that combines search terms as sets: AND requires all of them, OR accepts any of them and NOT excludes. → Unit 2

**Truncation**
The use of a symbol, usually the asterisk, to retrieve all the endings of a stem. → Unit 2

**Field searching**
A search limited to one part of the record, such as the title, the author or the abstract. → Unit 2

**Precision**
The proportion of retrieved results that are relevant. → Unit 2

**Recall**
The proportion of all relevant documents that the search retrieves. → Unit 2

**Berrypicking**
Marcia Bates's model in which searching is a process that changes as you read, picking up useful information at each step. → Unit 2

**Controlled vocabulary**
A set of standardised terms (thesaurus, subject headings) that a database uses to describe documents and group synonyms. → Unit 2

## Ethical use of information and plagiarism

**Intellectual property**
The set of authors' rights over their original creations: moral rights (which cannot be waived) and economic rights. → Unit 3

**Right of quotation**
A limit to intellectual property that allows fragments of other people's works to be included for analysis or commentary for teaching or research purposes, stating the source and the author. → Unit 3

**Public domain**
The status of a work whose economic rights have expired and which can be used freely, respecting moral rights. → Unit 3

**Creative Commons licence**
A licence with which the author authorises certain uses of their work in advance, combining the conditions BY, SA, NC and ND. → Unit 3

**Plagiarism**
Presenting someone else's work, ideas or words as your own without acknowledging them. → Unit 3

**Paraphrase**
Expressing an author's ideas in your own words and structure, accompanied by a citation. → Unit 3

**Patchwriting**
Copying a text while changing some words or their order, but keeping its structure. It is a form of plagiarism. → Unit 3

**Academic integrity**
A commitment to honesty, trust, fairness, respect and responsibility in study and research. → Unit 3

## References

**Citation**
A short marker in the text that points to the source used, such as (Bates, 1989) or [3]. → Units 4 and 5

**Reference**
The full description of a source in the final list of a piece of work. → Units 4 and 5

**Citation style**
A set of rules establishing which details citations and references include, in what order and with what punctuation (APA, Vancouver, ISO 690…). → Units 4 and 5

**Author-date system**
A citation system that identifies the source in the text by surname and year, and orders the list alphabetically. → Units 4 and 5

**Numeric system**
A citation system that identifies the source with a number and orders the list by order of appearance. → Units 4 and 5

**DOI**
*Digital Object Identifier*. A permanent identifier for a digital document that resolves at `https://doi.org/`. → Units 4 and 5

**Reference manager**
Software that stores and organises references, inserts citations into the word processor and generates the bibliography in a style. → Units 4 and 5

**Similarity-checking tool**
Software that compares a text with large collections of documents and flags overlaps. Its report is an indication, not a verdict of plagiarism. → Units 4 and 5

## Generative artificial intelligence

**Generative artificial intelligence**
AI that produces new content (text, images, audio, code) from patterns learned from large amounts of data. → Units 6 and 7

**Language model**
A system trained on large collections of texts to predict the most likely fragment that follows a given text. → Units 6 and 7

**Prompt**
The text with which you ask a generative AI system for something. → Units 6 and 7

**Hallucination**
An AI answer that looks correct but is false or invented, such as a non-existent reference. → Units 6 and 7

**Training data**
The collection of data with which a model is trained; it determines what the model knows and the biases it reproduces. → Units 6 and 7

**Retrieval-augmented generation**
*RAG*. A technique that combines a language model with a search engine or document collection to answer from what is retrieved. → Units 6 and 7

**Algorithmic bias**
A systematic error in an automated system that favours or disadvantages certain groups, often inherited from the data. → Units 8 and 9

**AI use statement**
A text stating which AI tool was used in a piece of work, what for, in which parts and how the result was checked. → Units 8 and 9

**AI Act**
Regulation (EU) 2024/1689, which regulates AI in the European Union with a risk-based approach. → Units 8 and 9

**AI literacy**
The knowledge and skills needed to use AI in an informed way and understand its opportunities, risks and possible harms. ⟶ Units 8 and 9

**Explicability**
An ethical principle according to which it must be possible to understand how an AI system works and who is accountable for it. ⟶ Units 8 and 9

**Personal data**
Any information about an identified or identifiable natural person. ⟶ Units 8 and 9

**Informational self-determination**
Each person's right to decide on the use of their personal data. ⟶ Units 8 and 9

# References

# Bibliography


American Psychological Association. *APA Style*. Retrieved https://apastyle.apa.org/

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Review*, *13*(5), 407–424. https://doi.org/10.1108/eb024320

Bawden, D., & Robinson, L. (2009). The dark side of information: Overload, anxiety and other paradoxes and pathologies. *Journal of Information Science*, *35*(2), 180–191. https://doi.org/10.1177/0165551508095781

Bender, E. M., & others. (2021,). *On the dangers of stochastic parrots*. https://doi.org/10.1145/3442188.3445922

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of FAccT '21: Proceedings of FAccT '21* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Creative Commons. *Licenses*. What each licence allows. Retrieved https://creativecommons.org/licenses/

Cruz Gil, C. (2015). *Modelos de búsqueda y recuperación de la información*. Trea.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.8cd550d1

Foltýnek, T., & others. (2020). Testing of support tools for plagiarism detection. *International Journal of Educational Technology in Higher Education*, *17*, 46. https://doi.org/10.1186/s41239-020-00192-4

Gilmour, R., & Cobus-Kuo, L. (2011). Reference management software: A comparative analysis of four products. *Issues in Science and Technology Librarianship*, *66*. https://doi.org/10.5062/F4Z60KZF

Kasneci, E., & others. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, *49*(4), 41–46. https://doi.org/10.1145/1121949.1121979

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Pacios Lozano, A. R. (2013). *Técnicas de búsqueda y uso de la información*. Ramón Areces.

Pecorari, D., & Petrić, B. (2014). Plagiarism in second-language writing. *Language Teaching*, *47*(3), 269–302. https://doi.org/10.1017/S0261444814000056

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, *25*(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Sugimoto, C. R., Work, S., Larivière, V., & Haustein, S. (2017). Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology*, *68*(9), 2037–2062. https://doi.org/10.1002/asi.23833

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*, *13*, 14045. https://doi.org/10.1038/s41598-023-41032-5