



Copyright & Fair Use

A Complete 10-Chapter Ethics Course

Topics covered:

What copyright protects · duration & the public domain · exclusive rights & infringement
The four-factor fair use test, in depth · fair use in real scenarios · Creative Commons licensing
DMCA takedowns · copyright in the age of AI · web scraping & copyright

Closing loops: this site's own L7 song-lesson rule, imgai1-9's AI-training note,
and dsproj1-1's scraping warn-box — each finally explained in full

Exercises: 30 hands-on scenarios with worked solutions

Format: A4 · Dark-theme code examples

General legal literacy — not legal advice.

Philip Osztromok · Generated with Claude

Table of Contents

- 01 What Copyright Actually Protects
- 02 Duration & the Public Domain
- 03 Exclusive Rights & Infringement
- 04 The Four-Factor Fair Use Test, In Depth
- 05 Fair Use in Practice — Real Scenarios
- 06 Creative Commons & Open Licensing
- 07 DMCA & Takedown Procedures
- 08 Copyright in the Age of AI
- 09 Web Scraping, Data Collection & Copyright
- 10 Capstone: A Real-World Copyright Decision Framework

CHAPTER 1 OF 10

What Copyright Actually Protects

Copyright & Fair Use

Chapter 1 · What Copyright Actually Protects

This course exists to close three loops the site has already opened without ever finishing: **L7**'s own permanent rule about checking whether song lyrics are copyright-free before generating a lesson, **imgai1-9**'s brief mention of AI-training-data copyright questions, and **dsproj1-1**'s own scraping-legality warn-box. Each of those moments assumed a working understanding of copyright without ever building one. This chapter starts building it — not with a list of rules to memorize, but with the one distinction almost everything else in this course depends on.

Copyright Is Automatic — No Registration, No Symbol Required

Copyright doesn't have to be applied for, registered, or marked with a © symbol to exist. In most modern copyright systems, it attaches automatically the moment an original work is **fixed in a tangible medium** — written down, recorded, saved to a file, painted on canvas. The instant a sentence is typed and saved, it is, by default, copyrighted.

THEN WHY REGISTER OR USE THE © SYMBOL AT ALL?

Registration and the © symbol aren't what create the copyright, but they still matter practically — registration is often required before filing an infringement lawsuit in the U.S., and can affect what damages are available. The symbol itself is now largely a courtesy notice rather than a legal requirement, but neither is what makes a work copyrighted in the first place.

The Idea/Expression Dichotomy

This is the single most load-bearing distinction in this entire course, and later chapters — infringement in Chapter 3, the fair use test in Chapter 4 — depend directly on it: **copyright protects the specific, particular expression of an idea, not the idea itself.**

A CONCRETE EXAMPLE

"A story about a young wizard who discovers he's famous in a hidden magical world and attends a boarding school for wizards" is an *idea* — nobody owns it, and any author is free to write their own story built on that same premise.

The actual text of a specific published novel using that premise is the *expression* — that specific arrangement of words, sentences, and scenes is what copyright protects.

This is why two authors can independently write two different stories sharing a broad premise without either infringing the other — the premise itself was never anyone's property to infringe.

What Qualifies as Protectable Expression

Two requirements have to be met before copyright attaches to something at all: it has to be **original** (independently created, with at least some minimal spark of creativity — not copied from something else), and it has to be **fixed** in some tangible form, not merely an idea held in someone's head.

"ORIGINAL" DOESN'T MEAN "EFFORTFUL"

A common misconception is that enough labor or diligence alone earns copyright protection — sometimes called the "sweat of the brow" theory. It doesn't. A U.S. Supreme Court case, *Feist Publications v. Rural Telephone Service* (1991), directly rejected this: a phone book's alphabetically-arranged listings, despite real effort to compile, were found not copyrightable, because arranging facts alphabetically involves no creative judgment at all — only creative *expression*, not effort, is what the law actually protects.

What Doesn't Qualify — Even If It Took Real Work to Produce

NOT PROTECTABLE	WHY
Facts	Discovered, not created — a historical date, a phone number, a scientific measurement belongs to no one
Short phrases, titles, names, slogans	Too short/functional to contain meaningful creative expression — though these can sometimes be protected by trademark instead, a genuinely different legal mechanism covered only briefly in this course
Ideas, procedures, methods, systems	The idea/expression dichotomy directly at work — only a particular expression of a method is protectable, never the underlying method itself
U.S. federal government works	Works created by the U.S. federal government are generally excluded from copyright protection entirely by statute

The Merger Doctrine — When Idea and Expression Collapse Into One

Occasionally an idea can only be expressed in one way, or a very small number of ways. When that happens, courts apply the **merger doctrine**: the expression "merges" with the underlying idea and loses copyright

protection, specifically to prevent one person from effectively monopolizing an entire idea by copyrighting its only possible expression.

A CONCRETE EXAMPLE

A simple mathematical rule stated in plain language ("to convert Celsius to Fahrenheit, multiply by nine-fifths and add thirty-two") has very few ways it can be phrased without becoming awkward or incorrect. Because the idea and its expression are practically inseparable here, the expression isn't given copyright protection — otherwise the underlying method itself would be effectively locked up.

Hands-On Exercises

EXERCISE 1

A cookbook author publishes a recipe. Explain, using this chapter's own material, which parts of that recipe (if any) are protectable by copyright and which aren't.

 [View solution](#)

EXERCISE 2

A colleague says, "I didn't put a © symbol on my document, so it isn't copyrighted — anyone can use it freely." Explain what's wrong with this statement.

 [View solution](#)

EXERCISE 3

Explain the merger doctrine in your own words, and describe why it exists as an exception rather than simply being a natural consequence of the idea/expression dichotomy already covered in this chapter.

 [View solution](#)

CHAPTER 1 QUICK REFERENCE

- Copyright attaches automatically on fixation — no registration or © symbol required for it to exist
- **Idea/expression dichotomy** — copyright protects a particular expression of an idea, never the idea itself
- Protection requires **originality** (creative judgment, not just effort) and **fixation** in a tangible form
- Facts, short phrases/titles, ideas/methods, and U.S. federal government works are not protectable by copyright

- **Merger doctrine** — when an idea has only one (or very few) possible expressions, that expression loses protection to avoid monopolizing the idea itself
- **Next chapter:** Duration & the Public Domain

CHAPTER 2 OF 10

Duration & the Public Domain

Copyright & Fair Use

Chapter 2 · Duration & the Public Domain

Chapter 1 established that copyright attaches automatically and lasts a while — but "a while" isn't forever. This chapter covers exactly how long protection lasts, and what happens to a work once it runs out: the public domain, a genuinely different legal state from anything covered so far, where a work can be used by anyone, for any purpose, with no permission needed at all.

How Long Copyright Actually Lasts

Duration depends on who created the work and under what circumstances. In the U.S. (this course's primary focus, per its own closing scope note), the modern standard for an individually-authored work is **the life of the author plus 70 years**. Works made for hire, and anonymous or pseudonymous works, follow a different rule entirely, since there's no individual author's lifespan to measure from.

TYPE OF WORK	DURATION
Individually authored, created 1978 or later	Life of the author + 70 years
Work made for hire, anonymous, or pseudonymous	95 years from publication, or 120 years from creation — whichever is shorter
Older, pre-1978 works	Governed by a more complex, since-superseded set of rules — genuinely more involved than this course's own general-literacy scope covers in depth

TERMS HAVE BEEN EXTENDED BEFORE, MORE THAN ONCE

Copyright term length isn't fixed forever — U.S. law has extended it multiple times over the past century, most notably in 1976 and again in 1998. Nothing about the current numbers in this chapter's own table should be treated as permanently fixed either; duration is a matter of current statute, not an unchangeable constant.

The Public Domain — A Genuinely Different Legal State

Once copyright expires, a work enters the **public domain**: it has no copyright protection at all, and can be copied, adapted, sold, performed, or built upon by anyone, for any purpose, without asking permission or

paying anyone. This is not the same as having a very permissive license — a public domain work has no copyright owner left to grant or withhold permission from in the first place.

REAL, CURRENT PUBLIC DOMAIN EXAMPLES

Works published in the U.S. before 1929 are, at the time of writing, definitively in the public domain. Copyright term expiration is an ongoing, rolling process, not a one-time event — each new year, on January 1st (informally nicknamed "Public Domain Day" by copyright commentators), another year's worth of works crosses that same threshold as their own term finally runs out.

How Else a Work Can Enter the Public Domain

Term expiration is the most common path, but not the only one:

- **Government works** — per Chapter 1, U.S. federal government works are generally excluded from copyright from the moment they're created, entering the public domain immediately rather than after a term expires
- **Voluntary dedication** — a copyright holder can deliberately place their own work into the public domain before any term would naturally expire, most commonly today via a formal tool like the CC0 public domain dedication (covered properly, and kept clearly distinct from an ordinary Creative Commons license, in Chapter 6)
- **Historical failure to renew** — under an older U.S. system (no longer in effect for new works), copyright required an affirmative renewal at a fixed point, and works whose owners simply failed to file that renewal fell into the public domain as a result — mostly a historical curiosity for older works today, not a live mechanism for new ones

"OLD" AND "PUBLIC DOMAIN" AREN'T AUTOMATICALLY THE SAME THING

A well-known real example: most of Arthur Conan Doyle's Sherlock Holmes stories fell into the public domain decades ago — but a handful of his very last stories, published in the 1920s, remained separately protected for years afterward under their own, later term. A 2014 U.S. court case (*Klinger v. Conan Doyle Estate*) directly addressed this: character traits and story elements that only first appeared in those still-protected final stories could not be freely used, even though the character and the bulk of the stories were otherwise long since public domain. A single well-known character can, in practice, be partially public domain and partially still protected at the same time, depending on which specific work a given detail actually came from.

Public Domain vs. Creative Commons — Not the Same Mechanism

This chapter deliberately does not cover Creative Commons licensing — that's Chapter 6's own material, and the two are kept separate on purpose, because they are genuinely different legal mechanisms, not two names for the same thing.

PUBLIC DOMAIN (THIS CHAPTER)	CREATIVE COMMONS LICENSE (CHAPTER 6)
No copyright exists on the work at all	Copyright still exists — the holder is voluntarily granting some specific permissions while keeping others
No conditions of any kind — no attribution required, no restrictions	Conditions typically attached (most commonly, at minimum, requiring attribution to the original creator)
No one can ever "revoke" public domain status once term expiration or dedication has occurred	Governed by the specific terms of the license chosen, which vary by license type

Hands-On Exercises

EXERCISE 1

A corporate training video was produced as a work made for hire in 1990 and never publicly released until 2010. Explain, using this chapter's own table, how its copyright term would be calculated.

 [View solution](#)

EXERCISE 2

Explain why the Sherlock Holmes example shows that "this character is old, so it must all be public domain" is an oversimplification, using the specific legal mechanism this chapter describes.

 [View solution](#)

EXERCISE 3

A friend says a Creative Commons Attribution license and the public domain are "basically the same thing, since both let you use the work for free." Explain what's genuinely different between the two.

 [View solution](#)

CHAPTER 2 QUICK REFERENCE

- Individually authored works (created 1978+): life of the author + 70 years
- Works made for hire / anonymous / pseudonymous: 95 years from publication or 120 years from creation, whichever is shorter
- Copyright terms have been extended by statute before — the current numbers aren't permanently fixed
- The **public domain** means no copyright exists at all — usable by anyone, for any purpose, with no permission needed

- A work can reach the public domain via term expiration, government-work status, voluntary dedication (e.g. CC0), or historical non-renewal
- Public domain and Creative Commons licensing are genuinely different mechanisms — one has no copyright at all, the other is copyright still in force with some permissions granted
- **Next chapter:** Exclusive Rights & Infringement

CHAPTER 3 OF 10

Exclusive Rights & Infringement

Copyright & Fair Use

Chapter 3 · Exclusive Rights & Infringement

Chapters 1 and 2 covered what copyright protects and how long that protection lasts. This chapter covers what protection actually *does* — the specific bundle of rights it hands the copyright holder — and what it means to violate one of those rights without permission or a legal exception.

The Bundle of Exclusive Rights

Copyright isn't one single right — it's a bundle of several distinct exclusive rights, each of which the copyright holder alone controls (or can license to someone else) unless a specific legal exception applies.

RIGHT	WHAT IT COVERS
Reproduction	Making copies of the work — the most fundamental of the bundle
Distribution	Selling, renting, or otherwise distributing copies of the work to the public
Public performance	Performing the work publicly — a play, a song, a film screening
Public display	Displaying the work publicly — an image, a photograph, artwork
Derivative works	Creating adaptations built on the original — a translation, a sequel, a remix, a film based on a book

Each right can be licensed independently — a copyright holder might permit reproduction (e.g. printing physical copies) while withholding the right to create derivative works (e.g. an unauthorized sequel), since the rights don't have to be granted or withheld as one indivisible package.

What Actually Counts as Infringement

Infringement means exercising one of these exclusive rights without authorization and without a legal exception (fair use, covered in Chapters 4 and 5, is the most significant such exception). Proving it typically requires two elements: **ownership of a valid copyright**, and **copying of original, protected expression** — directly tying back to Chapter 1's own idea/expression dichotomy, since copying an unprotected idea alone is never infringement, only copying the protected expression is.

HOW COPYING IS ACTUALLY PROVEN

Direct observation of someone literally copying a work is rare. Courts commonly rely on circumstantial evidence instead: showing the accused party had **access** to the original work, combined with **substantial similarity** between the two works, is often enough to support an inference of copying — even without a witness to the act itself.

INTENT ISN'T REQUIRED FOR DIRECT INFRINGEMENT

A common misconception is that infringement requires knowing you were doing something wrong. It doesn't — direct infringement can occur even when the infringer genuinely didn't realize the work was protected, or believed in good faith they had permission. "Innocent infringement" is still infringement, though it can affect what remedies a court ultimately awards — a topic of remedies and damages this course's own general-literacy scope doesn't cover in depth.

Direct vs. Contributory Infringement

Not everyone found liable for infringement is the one who personally copied the work. U.S. copyright law recognizes more than one way to be legally responsible.

TYPE	WHAT IT REQUIRES
Direct infringement	The party itself personally exercises one of the exclusive rights without authorization — e.g. actually copying, distributing, or displaying the work
Contributory infringement	Knowingly inducing, causing, or materially contributing to someone else's direct infringement — without necessarily infringing anything directly oneself
Vicarious infringement	Having the right and practical ability to control the infringing activity, and receiving a direct financial benefit from it — notably, this doesn't require actual knowledge the way contributory infringement does

A REAL, WELL-KNOWN EXAMPLE

Peer-to-peer file-sharing services in the early 2000s (Napster, and later Grokster) were found liable not because the companies themselves copied copyrighted songs, but because U.S. courts found they knowingly built and promoted a service whose primary, foreseeable use was enabling widespread infringement by their users — a real-world application of contributory (and, depending on the specific case, vicarious) infringement rather than direct infringement by the platform itself.

Hands-On Exercises

EXERCISE 1

A publisher licenses an author's novel for physical print distribution only, with no rights to film or translation. A studio later produces an unauthorized film adaptation. Explain, using this chapter's own bundle of rights, exactly which right

was violated and why the print license doesn't cover it.

 [View solution](#)

EXERCISE 2

Someone genuinely believes an image they found online was free to use and posts it on their own commercial website. It turns out to be copyrighted. Explain whether this is still infringement, and why.

 [View solution](#)

EXERCISE 3

Explain the difference between contributory and vicarious infringement using this chapter's own definitions, and explain why the Napster/Grokster example illustrates a platform's liability rather than a user's direct infringement.

 [View solution](#)

CHAPTER 3 QUICK REFERENCE

- Copyright is a **bundle** of separate exclusive rights: reproduction, distribution, public performance, public display, and derivative works — each licensable independently
- Infringement requires ownership of a valid copyright plus copying of protected expression — copying an unprotected idea alone is never infringement
- Access + substantial similarity is common circumstantial proof of copying, since direct observation is rare
- Direct infringement doesn't require intent — "innocent infringement" is still infringement, though it can affect remedies
- **Contributory infringement** — knowingly inducing or materially contributing to someone else's infringement
- **Vicarious infringement** — the right/ability to control the infringing activity plus a direct financial benefit, without requiring actual knowledge
- **Next chapter:** The Four-Factor Fair Use Test, In Depth

CHAPTER 4 OF 10

The Four-Factor Fair Use Test, In Depth

Copyright & Fair Use

Chapter 4 · The Four-Factor Fair Use Test, In Depth

Chapter 3 established what infringement is. This chapter covers the single most important exception to it — fair use, the legal doctrine that permits certain uses of copyrighted work without the copyright holder's permission, despite technically exercising one of the exclusive rights from Chapter 3's own bundle. This is the course's own central chapter, and the one every later scenario chapter (5, 8, 9, 10) builds directly on.

Fair Use Is a Defense, Not a Pre-Cleared Right

Fair use isn't a permission slip checked in advance — it's an **affirmative defense** raised in response to an infringement claim. Someone doesn't "have" fair use the way they'd have a license; a court weighs whether a specific, already-completed use qualifies as fair use only once the question is actually raised, typically after a copyright holder has already alleged infringement. In U.S. law, this doctrine is codified in the Copyright Act itself, and directs courts to weigh four specific factors.

Factor 1 — Purpose and Character of the Use

This factor asks what the use is actually doing, not just whether money changed hands. Commercial use weighs somewhat against fair use, and nonprofit/educational use weighs somewhat in favor — but neither is decisive on its own, and both are commonly outweighed by the more influential question this factor actually turns on: is the use **transformative**?

TRANSFORMATIVE USE, AND WHERE THE CONCEPT COMES FROM

A use is transformative when it adds new meaning, new message, or new purpose to the original, rather than simply standing in for it. This concept became central to fair use analysis following *Campbell v. Acuff-Rose Music* (1994) — the case that found 2 Live Crew's parody of "Oh, Pretty Woman" could qualify as fair use specifically because it commented on and transformed the original rather than merely reproducing it for the same purpose.

TRANSFORMATIVE USE HAS REAL LIMITS, CONFIRMED RECENTLY

"Just make it transformative enough" is not a reliable shortcut. In *Andy Warhol Foundation v. Goldsmith* (2023), the U.S. Supreme Court found that licensing a Warhol image derived from a photograph, for a magazine cover, was *not* fair use — because the new use served essentially the same commercial licensing purpose as the original photograph, despite Warhol's work being visually and artistically transformed. Visual or stylistic transformation alone doesn't

guarantee this factor favors fair use if the new use still competes with the original in the same market and for the same purpose.

Factor 2 — Nature of the Copyrighted Work

This factor looks at what kind of work is being used. Using a factual or informational work (news reporting, technical documentation) more readily supports fair use than using a highly creative, expressive work (a novel, a song, a film) — factual content sits closer to the unprotected "idea" side of Chapter 1's own idea/expression dichotomy, even where some original expression is still present. Using an **unpublished** work also weighs more heavily against fair use than using a published one, since an author's right to control the first public appearance of their own work carries independent weight.

Factor 3 — Amount and Substantiality of the Portion Used

This factor isn't purely about quantity — it's about both how much was used and how significant that portion is to the work as a whole.

QUALITY CAN MATTER MORE THAN QUANTITY

In *Harper & Row v. Nation Enterprises* (1985), a magazine published only a few hundred words from a soon-to-be-released memoir — a small fraction of the whole book. The U.S. Supreme Court still found this factor weighed against fair use, because the excerpt taken was the "heart of the work" — its most newsworthy, most memorable passage — showing that a small amount can still weigh heavily against fair use if it's the part that matters most.

Using only as much of a work as is genuinely necessary for the specific transformative purpose (a short clip for commentary, a brief quote for criticism) supports this factor; using more than that necessary amount, or the work's single most valuable portion, weighs against it.

Factor 4 — Effect on the Market

This factor asks whether the use harms the actual or potential market for the original work — including licensing markets the copyright holder might otherwise have been able to sell into. Historically treated by some courts as the single most important factor, though its relative weight compared to Factor 1's transformative-use analysis has shifted across different eras of fair use case law.

MARKET HARM MEANS MARKET SUBSTITUTION, NOT HURT FEELINGS

A harsh review or critical commentary that damages a work's reputation — and, as a result, its sales — is not the kind of "market harm" this factor is concerned with. The factor asks whether the new use *substitutes* for the original in the

marketplace (someone buys the copy instead of the original, or a licensing opportunity is lost), not whether the original creator dislikes the resulting commentary or feels reputationally harmed by legitimate criticism.

The Honest Core of This Whole Chapter

THIS IS A BALANCING TEST, NOT A CHECKLIST

No single factor is individually decisive, and courts weigh all four together, case by case, based on the specific facts in front of them — there is no fixed formula, point system, or combination of factors guaranteed to produce a particular outcome. Two uses that look superficially similar can be decided differently depending on details that seem minor on the surface. Nothing in this chapter — or this course — can substitute for an actual legal judgment on a specific real use; this chapter explains what courts weigh, not what any specific use will be decided to be.

A Worked Example, All Four Factors Together

A FILM REVIEW VIDEO USING SHORT CLIPS

- **Factor 1:** Likely favors fair use — critical commentary genuinely transforms the clips' original entertainment purpose into analysis and criticism, a different purpose than the film itself serves
- **Factor 2:** Weighs somewhat against — a film is highly creative, expressive work, not factual content
- **Factor 3:** Depends heavily on execution — short clips genuinely necessary to illustrate specific points support fair use; using extended, plot-revealing sequences or the film's single most pivotal scene weighs against it
- **Factor 4:** Likely favors fair use if the review doesn't substitute for watching the film itself — a review isn't generally a market substitute for the film, though a "clips reel" stitched together with minimal commentary could start to look like one

No single factor here is dispositive on its own — a court would weigh all four together, and a genuinely borderline case (long clips, minimal added commentary) could tip the outcome either way, exactly as the warn-box above describes.

Hands-On Exercises

EXERCISE 1

Explain why the Warhol/Goldsmith case complicates a simple "if it's transformative, it's fair use" rule, using this chapter's own material.

 [View solution](#)

EXERCISE 2

Explain why the Harper & Row case shows that Factor 3 isn't purely about how much of a work was used, using this chapter's own "heart of the work" reasoning.

 [View solution](#)

EXERCISE 3

A content creator argues that because their four-factor analysis on paper "adds up in their favor," their use is definitely fair use with no risk at all. Explain what's wrong with this reasoning, using this chapter's own central warning.

 [View solution](#)

CHAPTER 4 QUICK REFERENCE

- Fair use is an **affirmative defense**, decided after the fact by a court — not a pre-cleared right
- **Factor 1** — purpose/character, with transformative use as the most influential (but not unlimited) consideration
- **Factor 2** — nature of the work; factual/published content favors fair use more than creative/unpublished content
- **Factor 3** — amount used, both quantity and qualitative significance ("the heart of the work")
- **Factor 4** — market effect, meaning market substitution, not reputational harm from legitimate criticism
- This is a **case-by-case balancing test** — no factor is individually decisive, and no combination of factors guarantees an outcome
- **Next chapter:** Fair Use in Practice — Real Scenarios

CHAPTER 5 OF 10

Fair Use in Practice — Real Scenarios

Copyright & Fair Use

Chapter 5 · Fair Use in Practice — Real Scenarios

Chapter 4 laid out the four-factor test in the abstract. This chapter applies it to the categories of use people actually ask about most — commentary, parody, education, and news reporting — before closing with a direct, concrete walkthrough of a question this site itself has faced before: this course's own **L7** rule governing song lessons.

Commentary & Criticism

Quoting a passage from a book to critique it, or excerpting a scene from a film to analyze it, is one of the strongest and most traditionally recognized categories of fair use. Factor 1 favors it directly — criticism adds a new purpose (evaluation, analysis) genuinely distinct from the original's own purpose. It isn't unconditional, though: Chapter 4's own worked example already showed this can tip against fair use if the excerpt used is longer than necessary, or reaches into the work's own "heart" per *Harper & Row*.

Parody — And Why It's Legally Distinct From Satire

Parody specifically targets and comments on the *original work itself* — mimicking its style specifically to critique or mock that work. This is why *Campbell v. Acuff-Rose*'s parody of "Oh, Pretty Woman" (introduced in Chapter 4) leaned so heavily in favor of fair use: the parody needed to evoke the original specifically in order to comment on it.

SATIRE IS TREATED DIFFERENTLY, AND COURTS HAVE SAID SO DIRECTLY

Satire uses a work's style or elements to comment on something *else* entirely — not the original work itself. Because satire doesn't need to borrow from that particular work specifically (unlike parody, which inherently must reference its target to make its point), courts have been notably less willing to extend fair use to it. A 2020 federal appeals court case, *Dr. Seuss Enterprises v. ComicMix* (involving a Star Trek/Dr. Seuss mashup titled "Oh, the Places You'll Boldly Go!"), found the work was satire rather than parody — using Dr. Seuss's distinctive style to tell an unrelated Star Trek story, not to comment on Dr. Seuss's own work — and denied fair use on that basis.

Education & Classroom Use

A widespread misconception treats "it's for a classroom" as an automatic, blanket exemption from copyright. It isn't. Nonprofit educational purpose is genuinely one favorable consideration under Factor 1 — but it's one input into the same four-factor balancing test from Chapter 4, not a separate rule that overrides the other three factors.

A GENUINELY DIFFERENT MECHANISM, EASY TO CONFLATE WITH FAIR USE

Separate, narrower statutory exemptions do exist specifically for education — such as Section 110(1)'s exemption for certain face-to-face classroom performances, and the TEACH Act's own provisions for specific distance-education uses. These are distinct legal mechanisms from fair use itself, with their own specific requirements and limits, not simply another name for "education is fair use." Confusing a narrow statutory classroom exemption with the broader, more flexible (but less certain) fair use doctrine is a real, common mistake.

News Reporting

Reporting on a newsworthy event sometimes requires quoting from, or showing a portion of, a copyrighted work that is itself the subject of the news — a leaked document, a controversial video clip, a public figure's own copyrighted statement. This use pattern often favors Factor 1 (informational, not merely reproductive purpose) and, where only the newsworthy portion is used, Factor 3 as well — but breaking news status doesn't override Factor 4's market-effect analysis if the outlet reproduces so much of the original that it substitutes for a licensed use of the same material.

Revisiting L7 — This Site's Own Song-Lesson Rule

This site's own **L7** rule permits a full, line-by-line song lesson — with translation — only for copyright-free lyrics, while allowing a vocabulary-only extraction (**songvocab:**) from any song regardless of copyright status. This chapter can now explain, concretely, exactly why that line was drawn where it was.

WHY A FULL LINE-BY-LINE LESSON NEEDS COPYRIGHT-CLEAR LYRICS

A structured lesson translating and reproducing a song's lyrics line by line implicates two of Chapter 3's own exclusive rights at once: **reproduction** (the original lyrics themselves) and **derivative works** (a translation is itself a derivative work, per Chapter 3). Running the four-factor test honestly: Factor 1 is genuinely uncertain — a language lesson has an educational purpose, but per Chapter 4's own Warhol/Goldsmith lesson, "educational" alone doesn't guarantee transformative-use status, especially where the lesson largely reproduces the lyrics' own content and sequence for consumption rather than commenting on or critiquing them. Factor 2 weighs against fair use (song lyrics are highly creative expression). Factor 3 weighs heavily against it — a full line-by-line lesson uses effectively the entire work, not a limited excerpt. Given that combination, treating this use as reliably fair use would be optimistic at best — which is exactly why **L7** takes the more conservative path of requiring the underlying lyrics to already be copyright-free, rather than leaning on an uncertain fair use argument at all.

WHY VOCABULARY-ONLY EXTRACTION IS DIFFERENT, EVEN FROM A COPYRIGHTED SONG

songvocab: pulls out individual, decontextualized words — not the lyrics' actual sequence or phrasing. This connects directly back to Chapter 1: individual words alone are not protectable expression in the first place (similar in spirit to the short-phrases exclusion from Chapter 1's own table), so there's a real argument this doesn't even reach copyrighted *expression* to begin with, separate from the fair use question entirely. Even setting that aside and treating it as a fair use question, Factor 3 favors this use strongly — an isolated word list uses none of the song's own creative arrangement, none of its "heart" in the *Harper & Row* sense, which is precisely why **L7** treats this as safe regardless of the underlying song's copyright status.

Hands-On Exercises

EXERCISE 1

Explain, using this chapter's own material, why the Dr. Seuss/Star Trek mashup was found to be satire rather than parody, and why that distinction mattered to the outcome.

 [View solution](#)

EXERCISE 2

A teacher assumes that because a use is "for the classroom," it's automatically exempt from copyright entirely. Explain what's wrong with this assumption, distinguishing the two genuinely different legal mechanisms this chapter covers.

 [View solution](#)

EXERCISE 3

Using this chapter's own four-factor walkthrough, explain why a full line-by-line translated song lesson is a meaningfully riskier fair use argument than a vocabulary-only extraction from the same song.

 [View solution](#)

CHAPTER 5 QUICK REFERENCE

- **Commentary/criticism** — one of the strongest fair use categories, still bounded by Factor 3's amount-used limits
- **Parody** targets the original work itself; **satire** uses a work's style to comment on something else — courts treat the two differently
- Educational use is one favorable factor, not a blanket exemption — genuinely separate statutory exemptions (Section 110(1), the TEACH Act) exist and shouldn't be conflated with fair use itself
- News reporting often favors Factors 1 and 3, but doesn't override Factor 4 if too much of the original is reproduced

- **L7's own line, explained:** a full translated lyric lesson implicates reproduction + derivative-works rights and uses the whole work — genuinely uncertain fair use, hence the conservative copyright-free requirement; a vocabulary extraction uses no protectable expression in any meaningful amount, hence it's treated as safe regardless of the song's copyright status
- **Next chapter:** Creative Commons & Open Licensing

CHAPTER 6 OF 10

Creative Commons & Open Licensing

Copyright & Fair Use

Chapter 6 · Creative Commons & Open Licensing

Chapter 2 deliberately kept the public domain and Creative Commons licensing separate, promising a proper treatment of licensing later. This is that chapter — a full look at how Creative Commons actually works, and a direct correction of the single most common misunderstanding about it: that a CC license is just another way of saying "public domain."

What Creative Commons Actually Is

Creative Commons is a nonprofit organization that publishes standardized, free license templates. A copyright holder who applies one isn't giving up their copyright — they're using it to pre-authorize certain specific uses, under certain specific conditions, without requiring anyone to ask permission individually each time. The copyright still exists and is still owned by the creator; a CC license is the creator exercising that ownership to grant some rights in advance while keeping others.

The Building-Block License Elements

Every real Creative Commons license (aside from CC0, covered separately below) is built from a combination of these elements:

ELEMENT	WHAT IT REQUIRES OR RESTRICTS
BY (Attribution)	Credit must be given to the original creator — present in every real CC license
SA (ShareAlike)	Any derivative work must be released under the same or a compatible license
NC (NonCommercial)	Only noncommercial use is permitted
ND (NoDerivatives)	Only exact, unmodified copies may be shared — no adaptations or remixes at all

The Six License Combinations

LICENSE	WHAT IT MEANS IN PRACTICE
CC BY	Most permissive real license — reuse, remix, and commercial use all allowed, credit required
CC BY-SA	As above, but any derivative must carry the same license — the license Wikipedia itself uses
CC BY-NC	Credit required, commercial use prohibited
CC BY-NC-SA	Credit required, noncommercial only, derivatives must carry the same license
CC BY-ND	Credit required, exact copies only — no adaptations permitted
CC BY-NC-ND	The most restrictive combination — credit required, noncommercial only, exact copies only

CC0 — The Genuine Outlier

CC0 isn't really a "license" in the same sense as the six combinations above — it's a public domain dedication tool. A creator using CC0 waives all their copyright to the maximum extent legally possible, functionally placing the work into the public domain voluntarily, exactly as Chapter 2 previewed. Unlike every other CC option, CC0 requires no attribution and imposes no conditions at all — because, unlike the other six, it isn't granting conditional permissions under a retained copyright; it's giving the copyright up entirely.

The Core Myth, Corrected Directly

"CC-LICENSED" IS NOT "PUBLIC DOMAIN" — EVEN FOR THE MOST PERMISSIVE REAL LICENSE

Per Chapter 2's own comparison table, a public domain work has no copyright at all, with no conditions of any kind. Even the most permissive genuine Creative Commons license, CC BY, still legally requires attribution as a binding condition. Fail to give credit under a CC BY license and the use can fall outside the license's own terms — meaning it's no longer authorized use at all, but potential infringement of a copyright the creator never actually gave up. A public domain work carries no equivalent risk, because there's no license, and no retained copyright, left to violate.

Real Gotchas Worth Knowing

"NONCOMMERCIAL" IS GENUINELY AMBIGUOUS

What counts as "commercial use" under an NC license is a real, documented area of dispute — does a monetized video merely containing ads alongside NC-licensed material count as commercial use of that material, even if the creator never directly sells the licensed content itself? Different reasonable people, and different platforms' own guidelines, have answered this differently. Treat NC as a genuine gray area requiring judgment, not a bright line.

SHAREALIKE CAN MEANINGFULLY CONSTRAIN AN ENTIRE PROJECT

Incorporating SA-licensed material into a larger new work generally requires the *entire* resulting work to carry a compatible SA license — not just the specific portion that was reused. This "copyleft" effect can be a much bigger commitment than it first appears, especially for a project intended to be licensed differently overall.

A CLAIMED LICENSE ISN'T AUTOMATICALLY A VALID ONE

A CC license label attached to content online is only as reliable as whoever applied it — someone can mislabel content they never actually had the rights to license in the first place. Finding a "CC BY" tag on an image doesn't independently confirm the person who posted it actually held the copyright to grant that license. This is a real, practical risk worth remembering rather than blind trust in a label alone.

Hands-On Exercises

EXERCISE 1

Someone reuses a CC BY-licensed photo commercially without any attribution at all, assuming "CC just means free to use." Explain, using this chapter's own material, why this could still be copyright infringement.

 [View solution](#)

EXERCISE 2

Explain why CC0 is described as "the genuine outlier" among Creative Commons options, distinguishing it structurally from the other six license combinations.

 [View solution](#)

EXERCISE 3

A developer wants to incorporate a small CC BY-SA-licensed code snippet into a much larger project they intend to license restrictively. Explain the real consequence of doing so, using this chapter's own ShareAlike material.

 [View solution](#)

CHAPTER 6 QUICK REFERENCE

- A CC license doesn't give up copyright — the creator retains it while pre-authorizing specific uses under specific conditions
- **BY / SA / NC / ND** — the four building-block elements combining into six real license variants
- **CC0** is a genuine outlier — a public domain dedication, not a conditional license, with no attribution requirement at all

- **Core myth corrected:** even the most permissive real CC license (CC BY) still requires attribution — unlike true public domain, which has no conditions at all
- "Noncommercial" is genuinely ambiguous, ShareAlike can obligate an entire larger project, and a CC label isn't proof the labeler actually held the rights to apply it
- **Next chapter:** DMCA & Takedown Procedures

CHAPTER 7 OF 10

DMCA & Takedown Procedures

Copyright & Fair Use

Chapter 7 · DMCA & Takedown Procedures

Chapters 3 through 6 covered what rights exist, when they're infringed, and when a use is permitted anyway. This chapter covers the actual mechanism most infringement gets addressed through in practice on the modern web — not a lawsuit, but a fast, largely automatic notice-and-takedown process created by U.S. law in 1998.

What the DMCA Actually Set Up

The Digital Millennium Copyright Act's safe harbor provisions solved a real problem platforms faced in the early web: without some protection, a service hosting enormous volumes of user-uploaded content could face direct legal liability for infringing material its users posted, even without the platform's own knowledge or involvement. The DMCA created a **safe harbor** — legal protection from that liability — for platforms that follow a specific set of rules.

Safe Harbor's Real Requirements

A platform doesn't get safe harbor automatically just by existing — it has to actively qualify, by:

- Designating a DMCA agent to receive takedown notices
- Not having actual knowledge of specific infringing material (or acting quickly to remove it once made aware)
- Not receiving a direct financial benefit from infringement it has the right and ability to control — the same underlying idea as Chapter 3's own vicarious infringement standard, applied here to platforms specifically
- Implementing a policy for terminating repeat infringers' accounts

WHY THIS MATTERS FOR HOW THE INTERNET ACTUALLY WORKS

Safe harbor is a major reason platforms hosting massive amounts of user content (video sites, social media, forums) can exist at all without being sued out of existence over every piece of infringing content a user might post — but that protection is conditional, not unconditional, on actually running the notice-and-takedown process this chapter covers.

The Notice-and-Takedown Process, Step by Step

1. A copyright holder identifies infringing content hosted on a platform
2. They send a formal DMCA takedown notice to the platform's designated agent, including: identification of the copyrighted work, identification of the specific infringing material/URL, contact information, a good-faith statement that the use is unauthorized, and a statement made under penalty of perjury that the notice is accurate and the sender is authorized to act on the copyright holder's behalf
3. The platform removes or disables access to the content — acting "expeditiously" is a real legal requirement for maintaining safe harbor
4. The platform typically notifies the person who uploaded the content that it was removed and why

REMOVAL HAPPENS FIRST, BEFORE ANY JUDICIAL REVIEW

Nothing in this process requires a court to actually confirm infringement before content comes down. A takedown notice is a private party's own sworn claim — the platform removes the content to protect its own safe harbor status, not because a judge has ruled on the underlying copyright question. This single fact is the root of most of the real problems covered later in this chapter.

Counter-Notices

Someone whose content was removed can file a **counter-notice** if they believe the takedown was mistaken or the material was misidentified — asserting, again under penalty of perjury, that the removal was an error, and consenting to the jurisdiction of a relevant court.

WHAT HAPPENS AFTER A VALID COUNTER-NOTICE

If the platform receives a proper counter-notice, it must restore the material within a set window (commonly cited as 10–14 business days) *unless* the original copyright holder files an actual lawsuit before that window closes. In practice, this puts real pressure back on the copyright holder to either follow through with formal legal action or let the content be restored.

An Honest Look at Real Takedown Abuse

Because removal happens immediately, based on nothing more than a private party's own sworn claim, the system has a well-documented history of misuse.

A REAL, RECURRING PATTERN

False or bad-faith takedown notices have been used to silence legitimate criticism, suppress content from competitors, or remove genuine fair use material — precisely because removal is immediate while disputing it (via counter-notice, and potentially real legal exposure if the dispute escalates) takes real time, effort, and risk on the part of the person

whose content was taken down. Many uploaders simply don't contest a takedown at all, even when they'd likely have a strong fair use defense, exactly because of that asymmetry.

THE LAW DOES ADDRESS FALSE CLAIMS — ENFORCEMENT IS ANOTHER MATTER

Section 512(f) of the DMCA does create a legal claim against someone who knowingly files a materially false takedown notice. In practice, this provision is honestly inconsistent in enforcement — proving a false notice was made "knowingly" is a real evidentiary hurdle, and pursuing that claim requires the wrongly-targeted party to initiate their own separate legal action. The gap between what the law technically allows and what actually happens in practice is real and worth naming directly rather than glossing over.

This overall dynamic — where the mere threat or cost of a takedown can suppress legitimate speech or fair use even when the underlying copyright claim would likely fail — is often referred to as a "chilling effect," and is one of the most commonly cited criticisms of the notice-and-takedown system as it actually operates.

Hands-On Exercises

EXERCISE 1

Explain why a platform's safe harbor protection is described as "conditional, not unconditional," using this chapter's own list of requirements.

 [View solution](#)

EXERCISE 2

A creator's clearly fair-use commentary video gets taken down by a false DMCA claim. Explain, using this chapter's own material, why the creator might reasonably choose not to file a counter-notice even if they believe they'd win.

 [View solution](#)

EXERCISE 3

Explain the "chilling effect" this chapter describes, and why it can suppress legitimate fair use even when the underlying copyright claim would likely fail if actually tested in court.

 [View solution](#)

CHAPTER 7 QUICK REFERENCE

- **Safe harbor** — conditional protection from liability for platforms that follow the DMCA's own required process
- Notice-and-takedown: a sworn notice → expeditious removal → uploader notified — no judicial review required before content comes down
- **Counter-notice** — the uploader can dispute a takedown; the platform must restore the content unless the copyright holder actually sues within the response window
- False or bad-faith takedowns are a real, documented problem — removal is immediate while disputing it carries real cost and risk
- Section 512(f) technically penalizes knowingly false notices, but enforcement is honestly inconsistent in practice
- The **chilling effect** — the threat/cost of a takedown alone can suppress legitimate speech even when the underlying claim would likely fail
- **Next chapter:** Copyright in the Age of AI

CHAPTER 8 OF 10

Copyright in the Age of AI

Copyright & Fair Use

Chapter 8 · Copyright in the Age of AI

`imgai1-9` named training-data copyright and style mimicry as real issues without covering the legal mechanics underneath either one. This chapter provides that depth — and immediately splits what's often treated as one single question into three genuinely separate legal questions, each with its own, differently unsettled answer.

Three Separate Questions, Often Conflated Into One

QUESTION	WHAT IT ACTUALLY ASKS
A — The input question	Does training an AI model on copyrighted works itself infringe copyright?
B — The output-ownership question	Can AI-generated output itself be copyrighted?
C — The output-infringement question	Can AI-generated output itself infringe an existing copyrighted work?

These have different answers, different levels of legal settledness, and different reasoning behind each one — treating "AI and copyright" as a single question is itself a source of real confusion.

Question A — Is Training on Copyrighted Works Infringement?

This is genuinely, actively unsettled, and is the subject of multiple real, ongoing lawsuits at the time of writing. The core tension maps directly onto Chapter 4's own four-factor test:

ARGUMENT FOR FAIR USE	ARGUMENT AGAINST
Training extracts statistical patterns rather than storing retrievable copies — arguably a genuinely new, transformative purpose (Factor 1) distinct from the works' own original purpose	Training typically uses the entirety of each work (Factor 3), often for commercial purposes, and can plausibly compete with the market for the original works or their licensing (Factor 4)

A REAL, NOTABLE DATA POINT — NOT A FINAL ANSWER

In 2025, a U.S. federal court ruled in *Thomson Reuters v. Ross Intelligence* that using Westlaw's copyrighted legal headnotes to train a directly competing legal-research AI was *not* fair use — one of the first substantive rulings to

directly address AI training data. This involved a fairly specific fact pattern (training a direct market competitor on the plaintiff's own proprietary content), and shouldn't be read as settling the much broader question of AI training generally, especially against the backdrop of several other major cases (including disputes involving news publishers and stock-image libraries) still working through the courts.

Question B — Can AI-Generated Output Be Copyrighted?

Per current U.S. Copyright Office guidance, and confirmed by the courts in a case known as *Thaler v. Perlmutter* (which upheld the Office's refusal to register a fully AI-generated image with no claimed human author), the answer under existing U.S. law is: **not without sufficient human creative authorship**.

THIS CONNECTS DIRECTLY BACK TO CHAPTER 1

U.S. copyright law has long required a human author — a requirement that predates AI entirely but now sits squarely at the center of this question. Content generated purely by an AI system, with no meaningful human creative contribution, doesn't meet that requirement and isn't copyrightable at all — leaving it, functionally, without an owner in the way a human-authored work automatically has one per Chapter 1's own automatic-copyright rule.

This isn't absolute, though: where a human contributes sufficient, genuinely creative selection, arrangement, or editing on top of AI-generated raw material, that specific human contribution can itself be copyrightable — the Copyright Office has drawn a real, if genuinely fuzzy, distinction between merely typing a prompt (generally not enough on its own) and substantially selecting, arranging, or modifying AI output in a way that reflects real human creative judgment.

Question C — Can AI Output Itself Infringe an Existing Work?

Yes, in principle — and this is where Chapter 3's ordinary infringement analysis applies regardless of how the allegedly infringing material was actually produced. If a specific AI-generated output is substantially similar to a specific existing copyrighted work, the same analysis from Chapter 3 can apply exactly as if a human had copied it directly.

DOCUMENTED, REAL NEAR-REPRODUCTIONS

Some AI image generators have been documented producing outputs strikingly, sometimes near-exactly, similar to specific images that appeared in their own training data — undercutting the "it only learns statistical patterns, never reproduces anything specific" defense sometimes offered for Question A, at least for those specific documented cases.

STYLE MIMICRY IS LEGALLY DIFFERENT FROM REPRODUCING A SPECIFIC WORK

Per Chapter 1's own idea/expression dichotomy, a general artistic *style* functions much more like an unprotectable idea or method than protectable expression — nobody owns "painting in a loose impressionistic style" the way they own a specific painting. An AI generating new images "in the style of" a specific living artist, without reproducing any

specific protected expression from any of that artist's actual works, sits on genuinely uncertain legal ground under a straightforward Question C analysis — a real, serious ethical concern (exactly as `imgai1-9` named it) that doesn't automatically translate into a strong copyright infringement claim the way reproducing a specific existing work more directly would.

The Honest Bottom Line

THIS IS A LIVE, MOVING TARGET

AI and copyright is currently one of the most actively litigated, least settled areas of copyright law. Court rulings are arriving, but often address narrow fact patterns rather than resolving the broader questions definitively, and the law in this area will likely keep developing significantly. Treat everything in this chapter as a snapshot of where things currently stand, not a permanent, final answer — the same case-by-case caution Chapter 4 applied to fair use generally applies here with even less settled ground underneath it.

Hands-On Exercises

EXERCISE 1

Explain why treating "AI and copyright" as one single legal question is a mistake, using this chapter's own three-question breakdown.

 [View solution](#)

EXERCISE 2

Explain why a fully AI-generated image with no human creative input isn't copyrightable under current U.S. law, connecting your answer back to Chapter 1's own material.

 [View solution](#)

EXERCISE 3

Explain why "generating images in a living artist's style" is a real ethical concern per `imgai1-9` but a genuinely uncertain copyright infringement claim, using this chapter's own idea/expression reasoning.

 [View solution](#)

CHAPTER 8 QUICK REFERENCE

- Three separate legal questions: (A) is training infringement, (B) can AI output be copyrighted, (C) can AI output infringe an existing work
- **Question A** is genuinely unsettled — a real Factor 1 (transformative) vs. Factor 3/4 (full use, market effect) tension, with *Thomson Reuters v. Ross Intelligence* (2025) as a real but narrow data point
- **Question B** — under current U.S. law, AI output without sufficient human creative authorship isn't copyrightable at all (*Thaler v. Perlmutter*); meaningful human selection/arrangement/editing can change this
- **Question C** — ordinary infringement analysis applies regardless of how output was produced; documented near-exact reproductions of training images have occurred
- Style mimicry is a real ethical concern but legally different from reproducing a specific work — style itself is generally not protectable expression
- This entire area is actively litigated and unsettled — treat this chapter as a snapshot, not a final answer
- **Next chapter:** Web Scraping, Data Collection & Copyright

CHAPTER 9 OF 10

Web Scraping, Data Collection & Copyright

Copyright & Fair Use

Chapter 9 · Web Scraping, Data Collection & Copyright

`dsproj1-1` built a real scraper and flagged "check `robots.txt` /ToS" as good practice without ever explaining *why*. This chapter finally answers that — and, in the spirit of Chapter 8's own three-question breakdown, immediately splits "is scraping legal?" into several genuinely separate legal questions, of which copyright is only one.

"Is Scraping Legal?" Is Not One Question

CONSIDERATION	WHAT BODY OF LAW IT ACTUALLY FALLS UNDER
Is the scraped content copyrightable?	Copyright law — this course's own primary subject
Does using the site require agreeing to terms prohibiting scraping?	Contract law — genuinely separate from copyright
Does collecting the data require bypassing a technical access barrier?	Computer-fraud-type statutes (in the U.S., the CFAA) — also separate from copyright

A scraping project can be entirely fine on one of these dimensions and genuinely risky on another — treating "scraping" as governed by a single unified law is exactly the kind of oversimplification this course has repeatedly corrected elsewhere.

The Copyright Dimension

The technical act of scraping — copying content from a page into your own systems — does implicate the reproduction right from Chapter 3, *if* what's being copied is copyrightable expression. But a great deal of what data scientists actually scrape is exactly the kind of material Chapter 1 already excluded from protection entirely.

A DIRECT CALLBACK TO CHAPTER 1

Prices, stock levels, sports scores, weather readings, structured product specifications — bare facts, per Chapter 1's own table, "discovered, not created," and not protectable by copyright at all. Scraping this kind of factual data for

analysis generally doesn't implicate copyright in any meaningful way, regardless of what a site's own `robots.txt` or ToS might separately say about it.

CREATIVE/EXPRESSIVE CONTENT IS A DIFFERENT MATTER

Scraping actual articles, reviews, or blog posts — genuinely creative written expression, not bare facts — does implicate real copyrightable material. Whatever happens to that scraped text afterward (storing it, redistributing it, using it to train a model per Chapter 8's own Question A) needs its own separate fair use analysis under Chapter 4's four factors, exactly as if the text had been copied any other way.

Correcting a Common Myth: robots.txt Is Not a Law

A VOLUNTARY TECHNICAL SIGNAL, NOT A BINDING LEGAL BARRIER

`robots.txt` is a voluntary technical convention — a file a site publishes to signal which pages it would prefer automated crawlers avoid. It is not, by itself, a legally binding restriction in most circumstances, and ignoring it isn't automatically a copyright violation or any other independent legal claim on its own. That said, ignoring it is against widely accepted web norms, and disregarding it can be cited as evidence of bad faith in a separate legal dispute — it's a real signal worth respecting, just not because the file itself carries independent legal force the way a statute does.

Terms of Service — A Genuinely Different, Binding Mechanism

Unlike `robots.txt`, a site's Terms of Service can be a real, binding contract — if the scraper actually agreed to it (by creating an account, clicking "I agree," or in some cases simply by using a site whose terms are clearly presented). Violating a ToS provision that prohibits scraping is a potential **breach of contract**, an entirely different legal claim from copyright infringement, with its own separate remedies (account termination, contract damages) rather than the copyright-specific mechanisms covered in Chapters 3 and 7.

Bypassing Technical Access Controls — A Third, Separate Concern

Actively defeating a login wall, a paywall, or a CAPTCHA to reach data goes beyond ordinary scraping and can raise concerns under computer-fraud-type statutes — in the U.S., most notably the Computer Fraud and Abuse Act (CFAA) — which is a body of law entirely separate from copyright.

A REAL, SIGNIFICANT CASE: HIQ LABS V. LINKEDIN

In this closely watched U.S. case, the Ninth Circuit Court of Appeals found that scraping data that is *publicly accessible* — not hidden behind a login — generally does not violate the CFAA, since the statute concerns "unauthorized access," and data openly available without authentication isn't accessed without authorization in the relevant sense. The case had a genuinely complex procedural history (including being reconsidered after a separate U.S. Supreme Court decision, *Van Buren v. United States*, narrowed how courts interpret the CFAA) and eventually settled — a real, useful

data point, not a permanently settled rule for every scraping scenario. The general shape it supports, though, is a genuinely useful one: scraping *publicly accessible* data sits on meaningfully more solid legal ground under CFAA-type analysis than scraping data behind a login or other access barrier.

Putting It Together — Why dsproj1-1's Advice Was Right, For Reasons It Never Explained

GOOD PRACTICE REDUCES RISK ACROSS SEVERAL SEPARATE LEGAL DIMENSIONS AT ONCE

Checking a site's `robots.txt` and ToS before scraping isn't legally required in every single case, but it's genuinely good practice because it touches all three dimensions covered in this chapter simultaneously: it respects the voluntary technical signal (goodwill, and evidence against a bad-faith argument), it surfaces any binding contractual restriction before it's violated (the genuine ToS/contract-law risk), and checking whether data sits behind a login wall at all is the single clearest signal for whether CFAA-type concerns are even in play. `dsproj1-1`'s own advice was sound — this chapter is simply the legal reasoning underneath it that chapter never had room to cover.

Hands-On Exercises

EXERCISE 1

A data scientist scrapes only publicly listed product prices from a retail site, with no login required. Explain, using this chapter's own material, why this is very likely fine from a copyright standpoint specifically — while noting what it doesn't tell you about the other two considerations.

 [View solution](#)

EXERCISE 2

A colleague says, "our scraper ignores robots.txt, so we're definitely breaking the law." Explain what's wrong with this statement, and what robots.txt actually is instead.

 [View solution](#)

EXERCISE 3

Explain what the *hiQ Labs v. LinkedIn* case suggests about the difference between scraping publicly accessible data and scraping data behind a login wall, and why the case shouldn't be treated as a permanently settled rule.

 [View solution](#)

CHAPTER 9 QUICK REFERENCE

- "Is scraping legal?" splits into three separate questions: copyright, Terms of Service (contract law), and access-control circumvention (CFAA-type law)
- Bare factual data (prices, scores, structured specs) generally isn't copyrightable at all, per Chapter 1 — scraping it usually doesn't implicate copyright meaningfully
- Creative/expressive scraped text (articles, reviews) does implicate real copyright — any later use needs its own fair use analysis
- **robots.txt is not legally binding by itself** — a voluntary technical convention, not a law, though ignoring it can be cited as evidence of bad faith elsewhere
- A site's **Terms of Service** can be a genuinely binding contract if actually agreed to — violating it is breach of contract, not copyright infringement
- Bypassing login walls/CAPTCHAs raises CFAA-type concerns; *hiQ Labs v. LinkedIn* suggests publicly accessible data scraping sits on firmer ground than login-gated scraping, though the case has real, complex nuance
- **Next chapter:** Capstone — A Real-World Copyright Decision Framework

CHAPTER 10 OF 10

Capstone: A Real-World Copyright Decision Framework

Copyright & Fair Use

Chapter 10 · Capstone — A Real-World Copyright Decision Framework

Nine chapters have built the pieces: what's protected, how long, what infringement means, when fair use applies, how licensing and enforcement actually work, and how all of it plays out in two genuinely hard modern cases. This capstone assembles those pieces into one practical decision framework, and runs it against three realistic scenarios — closing the loop this whole course opened back in Chapter 1.

The Decision Framework

1. **Is the work actually protected by copyright at all?** — check whether it's protectable expression rather than an idea/fact (Chapter 1), and whether its copyright term has already expired into the public domain (Chapter 2)
2. **Does the intended use exercise one of the exclusive rights?** — reproduction, distribution, performance, display, or derivative works (Chapter 3)
3. **Is there already permission for this specific use?** — an explicit license (Creative Commons or otherwise, Chapter 6) or direct permission from the rights holder
4. **If not, does fair use plausibly apply?** — run the real four-factor analysis (Chapters 4–5), honestly, factor by factor — remembering it's a case-by-case balancing test, never a guarantee
5. **Are there separate legal considerations beyond copyright itself?** — Terms of Service, access-control circumvention, and other non-copyright issues (Chapter 9) that a "yes, this is fair use" answer alone would never address
6. **What's the actual real-world risk, given the stakes?** — even a strong fair use argument can trigger a takedown before anyone reviews it (Chapter 7), and "probably fine" and "risk-free" are not the same thing

Scenario A — Using a Melody in a Language Lesson

CASE 1: A TRADITIONAL FOLK MELODY

Step 1 — many traditional folk melodies are old enough to be in the public domain (Chapter 2); if confirmed, the analysis ends here — no permission or fair use argument is even needed, since there's no copyright left to implicate.

CASE 2: A SPECIFIC, RECENTLY COMPOSED MELODY

Step 1 — still protected. Step 2 — using it in a lesson implicates reproduction, and likely performance rights (Chapter 3). Step 3 — no license or permission exists. Step 4 — this closely mirrors Chapter 5's own worked analysis of L7's song-lesson rule: Factor 2 weighs against fair use (highly creative work), and depending on how much of the melody is used, Factor 3 could weigh against it too if the whole tune is reproduced. This is exactly the kind of case where L7's own conservative, copyright-clear-material-only approach is the right call rather than leaning on an uncertain fair use argument.

Scenario B — Scraping a Site to Build a Training Dataset

APPLYING CHAPTERS 8 AND 9 TOGETHER

Step 1 — depends entirely on what's scraped: bare factual data (Chapter 1) likely isn't protected at all; scraped articles or creative text are. Step 2 — building a training dataset involves reproduction (Chapter 3). Step 5 — this is where Chapter 9's own three-way split matters most: check the site's Terms of Service for a scraping prohibition (a separate contract-law question), and check whether the data required bypassing a login or other access control (a separate CFAA-type question) — neither of which the copyright analysis alone would ever surface. Step 4 — if the scraped content is creative and copyrighted, whether training on it is fair use is exactly Chapter 8's own Question A: genuinely unsettled, actively litigated, and not something this framework — or any single chapter of this course — can resolve with certainty.

Scenario C — Reusing an Image Found Online

THE SINGLE MOST COMMON REAL-WORLD SCENARIO

Step 1 — check whether the image is genuinely public domain (Chapter 2) rather than merely old-looking. Step 3 — check for an actual Creative Commons license on the source (Chapter 6), and if found, check which specific variant and whether its conditions (attribution, noncommercial-only, no-derivatives) are actually met for the intended use — remembering Chapter 6's own core myth-correction that a CC license is not the same as no restrictions at all. If neither public domain nor a valid license applies, Step 4 — run the four-factor test honestly for the specific intended use (a hero image on a commercial site is a very different Factor 1/Factor 4 analysis than a small thumbnail accompanying genuine critical commentary). Step 6 — remember Chapter 7's own lesson: even a strong fair use position doesn't prevent an initial takedown notice from a rights holder who disagrees.

Chapter Attribution

FRAMEWORK STEP	SOURCE CHAPTER(S)
1 — Is it protected at all?	copyright1-1, copyright1-2
2 — Which exclusive right is implicated?	copyright1-3
3 — Is there existing permission?	copyright1-6

FRAMEWORK STEP	SOURCE CHAPTER(S)
4 — Does fair use apply?	copyright1-4, copyright1-5
5 — Any separate legal considerations?	copyright1-9
6 — What's the real-world risk?	copyright1-7
Scenario B's AI/training dimension	copyright1-8

Honest Scope Note

THIS IS LEGAL LITERACY, NOT LEGAL ADVICE

- This course explains how copyright law generally works and gives you the vocabulary and framework to reason about it — it does not, and cannot, tell you with certainty how a specific real dispute would actually be decided
- This course is written primarily around U.S. law. Jurisdiction genuinely changes the analysis — for one concrete example, many Commonwealth countries (the U.K., Canada, Australia) use "fair dealing" instead of fair use: a narrower doctrine limited to a specific, enumerated list of permitted purposes (research, private study, criticism, review, news reporting), rather than the U.S.'s own open, flexible four-factor balancing test from Chapters 4–5. The two are genuinely not interchangeable, and assuming U.S.-style fair use reasoning applies everywhere is a real, common mistake
- A real decision with real stakes — a commercial project, a genuine legal dispute, anything with meaningful money or risk attached — should involve an actual lawyer, not just this course's own general framework

CLOSING THE LOOP THIS COURSE OPENED

This course exists to close three real loops: `L7`'s own song-lesson copyright rule, `imgai1-9`'s AI-training-data mention, and `dsproj1-1`'s scraping-legality warn-box. Each of those now has a real, worked explanation behind it rather than an unexplained rule — Chapter 5 for the first, Chapter 8 for the second, Chapter 9 for the third.

Hands-On Exercises

EXERCISE 1

Using this chapter's own six-step framework, walk through a scenario of your own choosing (a real or hypothetical reuse of copyrighted material) and identify which step, if any, would stop you from proceeding without further permission.

 [View solution](#)

EXERCISE 2

Explain why Scenario B (the training dataset) requires applying both Chapter 8's and Chapter 9's material together, rather than either chapter alone being sufficient.

 [View solution](#)

EXERCISE 3

Explain why "fair dealing" and "fair use" are described as genuinely not interchangeable, and why assuming otherwise is a real, common mistake.

 [View solution](#)

CHAPTER 10 QUICK REFERENCE — COURSE COMPLETE

- A six-step framework: protected? → which right? → existing permission? → fair use? → other legal considerations? → real-world risk?
- Public domain/CC-licensed material can end the analysis early (Scenario A, Case 1; Scenario C) — always check before reaching for fair use
- Scraping and AI training scenarios require checking copyright, contract law, and access-control law *separately* — no single check covers all three
- Jurisdiction genuinely matters — U.S. fair use and Commonwealth fair dealing are meaningfully different doctrines, not two names for the same thing
- This course is general literacy — real decisions with real stakes need an actual lawyer, not just this framework