Domain Knowledge

Written by [Febogamedev]

When verifying a response produced by a model, consulting external sources is often necessary (Autio et al., 2024; Ji et al., 2023).
Before searching indiscriminately for information, however, we need to understand which parts of the output truly deserve attention and which checks are most appropriate (Chi et al., 1981).
Domain knowledge serves this diagnostic function.

Having domain knowledge does not mean knowing every detail of a field. It means having at least an essential map consisting of (Chi et al., 1981; Ericsson & Lehmann, 1996):

- Fundamental definitions.
- Technical vocabulary.
- Typical relationships between concepts.
- Constraints and invariants.
- Plausible orders of magnitude.
- Key authoritative sources.
- Relevant edge cases and exceptions.

This map allows the output to be compared with structures independent of the quality of its presentation.
A text can be fluent, organized and convincing even when it contains simplifications, omissions or errors (Ji et al., 2023; Reber & Schwarz, 1999).
Without a minimum knowledge of the field, we therefore risk judging it mainly by how it is written rather than by the soundness of its content (Chi et al., 1981; Reber & Schwarz, 1999).

Example
Suppose an *LLM* produces this explanation:
“In the *HTTP* protocol, every router that forwards a packet sends an *ACK* to the previous router to confirm that the data was received correctly.”

The sentence is clearly written and may seem plausible to someone unfamiliar with networks.
However, a person with a minimum of domain knowledge already has an essential conceptual map:

- HTTP belongs to the application layer (Fielding et al., 2022).
- Routers operate primarily at the network layer (Deering & Hinden, 2017).
- ACKs are typically associated with protocols that include acknowledgment mechanisms, such as TCP (Eddy, 2022).
- IP forwarding between routers does not normally involve an end-to-end ACK for every packet received (Deering & Hinden, 2017; Eddy, 2022).
- Network layers serve different functions.

This knowledge is not yet enough to reconstruct every technical detail, but it is sufficient to reveal an anomaly:

HTTP → router → ACK

The relationship looks suspicious because it mixes concepts belonging to different layers and protocols.

At this point, domain knowledge serves a diagnostic function: it does not yet establish with certainty which part of the sentence is false, but it indicates where to focus verification. Instead of searching generically for “how the Internet works”, one can check specific questions:

- “Do IP routers send an ACK to the previous router for every packet?”.
- “To which layer do TCP ACKs belong?”.
- “Does HTTP directly manage packet delivery acknowledgments?”.

Verification can then show that the initial response confuses mechanisms belonging to different layers.

The process can be summarized as follows:
Plausible output → domain knowledge → anomaly recognized → targeted verification

Without that minimum knowledge, the reader might instead judge mainly the formal quality of the sentence:
“It is clear, uses technical terms and seems coherent, so it is probably correct”.

The central point, therefore, is that domain knowledge does not necessarily replace external sources, but helps us understand what to verify, where an error might lie and how plausible a claim is even before searching (Chi et al., 1981).

Domain knowledge alone does not demonstrate that a response is correct or incorrect. However, it allows us to recognize anomalies, identify critical claims and formulate more precise verification questions (Autio et al., 2024; Chi et al., 1981).
In summary, domain knowledge does not replace sources: it guides verification.

Domain knowledge helps distinguish three different situations.

- Obvious error: something incompatible with the field.
- Subtle error: something plausible but incomplete.
- Gray area: something that remains possible but requires further evidence.

Obvious error
An obvious error openly violates a fundamental rule, fact or relationship in the field.
It may appear as:

- An impossible chronology.
- An incorrect unit of measurement.
- An incompatible formula.
- A contradictory definition.
- A value outside any plausible order of magnitude.
- A technical term used with the wrong meaning.

For example, stating that a digital photograph measuring 1920 × 1080 pixels contains a total of 2,073 pixels would be an error easily recognized by someone who knows the domain.

Subtle error
A subtle error retains an appearance of plausibility but simplifies or distorts the problem.
It may involve:

- An overly narrow definition.
- A causal relationship presented as direct.
- An important exception ignored.
- A technical term used imprecisely.
- An attribution expressed with excessive confidence.
- A conceptual distinction flattened out.

These errors are harder to recognize because they do not necessarily produce obviously absurd sentences (Ji et al., 2023; Reber & Schwarz, 1999).
They often require sufficient expertise to understand that a step, condition or clarification is missing.

Gray area
The gray area includes claims that are plausible but insufficiently grounded (Ji et al., 2023).
In these cases, the output is not clearly false, but the available evidence is not robust enough to consider it reliable without further verification or qualification.

Example
"Using a cool palette makes a scene more melancholy".

This claim falls within the gray area.
The relationship may be plausible in some contexts, but not in every context.
The effect of colors also depends on other factors, such as composition, subject, contrast, cultural conventions and communicative intention (Jonauskaite et al., 2020).

A more cautious formulation would be:
"A cool palette can help create a melancholy atmosphere, but the result also depends on the other elements of the scene".

To use domain knowledge systematically, it is useful to construct a minimal model of the field under consideration (Chi et al., 1981).
This model should not contain everything we know.
It should collect the elements needed to evaluate the specific task.
It may include: basic definitions, fundamental relationships, orders of magnitude, necessary conditions, terminological distinctions, major exceptions and reference sources (Chi et al., 1981; Ericsson & Lehmann, 1996).

Example
Imagine having to evaluate a response about *JPEG* compression. The minimal model might include:

- The difference between lossy and lossless compression.
- The relationship between quality, file size and artifacts.
- The raster nature of the image.
- The role of transformation and quantization (International Telecommunication Union [ITU], 1992).
- The difference between resolution and compression quality.

This map allows us to quickly recognize claims such as:
“Reducing JPEG quality decreases the image’s resolution.”

The sentence seems plausible, but confuses two different properties: compression can degrade visual quality without necessarily changing the number of pixels (ITU, 1992).

An invariant is a relationship or condition that must remain valid within the problem under consideration.
Invariants can serve as quick tests of the output.

For example:
- A percentage must fall within a range compatible with its definition.
- A chronology must respect the temporal order of events.
- An effect cannot precede its stated cause.
- A physical quantity must be associated with the appropriate unit.
- A category cannot simultaneously be defined through incompatible criteria.
- A technical term should not change meaning without this being indicated.

If the output violates one of these constraints, there is a concrete reason to investigate further.
Invariants do not always demonstrate the correct solution, but they often allow us to rule out solutions incompatible with the structure of the domain (Hernán & Robins, 2020).
Many errors produced by models concern implausible quantitative values rather than complex concepts (Ji et al., 2023).

An order-of-magnitude sanity check consists in asking whether numbers, durations, percentages, distances or proportions are compatible with what normally occurs in the domain (Chi et al., 1981).

For example:
- Can a photographic file realistically be that size?
- Can a transfer speed produce the stated time?
- Can a percentage exceed the total being considered?
- Is the indicated cost plausible for the type of service?
- Is the date compatible with the chronology of events?
- Is the number of users consistent with the size of the market?

It is not always necessary to know the exact value.
Sometimes it is enough to recognize that the proposed value is too high, too low or incompatible with other data in the response.

A sanity check does not replace precise verification, but quickly identifies claims that deserve checking (Autio et al., 2024).

Every domain uses terms that define important distinctions.
In this text, we will call technical terms that keep the conceptual boundaries of the discussion stable **anchor words**.
For example, in a text about computer networks, anchor words may include: frame, packet, segment, *IP* address, *MAC* address, socket, port.

If the model uses these terms as synonyms, or changes their meaning over the course of the response, this may indicate that it is flattening important distinctions (Chi et al., 1981).
The same can happen in other fields.

Checking the stability of vocabulary allows us to recognize errors that emerge not from a single sentence, but from the relationship between different parts of the output.

Reconstructing the causal chain

One of the most useful applications of domain knowledge consists in reconstructing the causal chain (Hernán & Robins, 2020).
When the model states “X causes Y”, asking whether the sentence sounds plausible is not enough.
It is necessary to reconstruct the steps through which X is supposed to produce Y.

The fundamental question is: “Through which conditions and intermediate steps is X supposed to produce Y?”

Every domain has typical causal relationships between conditions, processes and effects. Compressing them excessively can produce fluent but epistemically weak explanations (Hernán & Robins, 2020; Hill, 1965).
A causal relationship should therefore be analyzed by considering:

- The intermediate mechanism.
- Necessary conditions.
- Contributing factors.
- Edge cases.
- Exceptions.
- Alternative causes.
- The difference between causality and mere correlation.

Example
Consider the claim:
“A sequence shot increases viewer immersion because it eliminates editing cuts.”

The sentence is plausible, but compresses several steps.

One possible causal chain could be:
absence or concealment of cuts → greater perceptual continuity → more stable construction of the scene's space and time → possible increase in viewer engagement.

This reconstruction makes the relationship more understandable, but also introduces conditions and limitations.
Not every cut reduces immersion.

- Coherent editing can be perceived without interrupting engagement.
- Likewise, a long take can become slow, contrived or disorienting.

The effect also depends on: staging, camera movement, acting, rhythm, composition, sound, genre and viewer expectations.

The initial claim is therefore not necessarily false, but it is causally incomplete.
Domain knowledge allows us to identify the missing steps and understand which parts require external evidence.

A well-constructed causal chain can be convincing and still be false (Hernán & Robins, 2020).

The model can generate an elaborate explanation in which every step seems plausible, even when sufficient empirical evidence is lacking.
Causal reconstruction must therefore be considered a test of internal consistency, not a demonstration (Hernán & Robins, 2020; Hill, 1965).
It serves to ask:

- Can the causal link be articulated?
- Are the steps compatible with the domain?
- Are important conditions missing?
- Are there alternative causes?
- Does the text confuse correlation and causality?
- Which steps require a source?

When the claim is significant, the reconstructed chain must be compared with studies, data, tests or independent sources (Hernán & Robins, 2020; Hill, 1965).
A coherent explanation can be plausible without being demonstrated.

To make a causal relationship explicit, targeted questions can be asked:

- What intermediate steps connect X to Y?
- What conditions must be present for X to produce Y?
- Is the relationship direct or mediated by other factors?
- In which cases does the relationship not hold?
- What alternative factors can produce the same effect?
- Are you describing causality or correlation?
- What evidence supports each step?

These questions do not guarantee that the model will correctly recognize its own errors.
However, they make visible the steps that the initial text left implicit (Huang et al., 2024).

- If the response remains vague or cannot coherently connect cause and effect, the claim requires more thorough checking.
- If it instead produces a plausible chain, the decisive steps will still need to be verified using independent sources or data (Autio et al., 2024; Huang et al., 2024).

Applying domain knowledge

Domain knowledge becomes truly useful when it is translated into specific checks.
Instead of asking: “Are you sure?”
it is more effective to ask:

- “What definition are you using?”.
- “What conditions must be true for this claim to hold?”.
- “What are the edge cases?”.
- “Are there counterexamples?”.
- “What unit of measurement was used?”.
- “What is the expected order of magnitude?”.
- “Does the term retain the same meaning throughout the response?”.
- “What steps connect the cause to the effect?”.
- “Which primary source does this data come from?”.
- “What alternative explanations are possible?”.

A generic request for confirmation often invites the model to reiterate the initial response (Huang et al., 2024).
By contrast, domain-specific questions define the point to analyze and make potential weaknesses easier to identify (Chi et al., 1981).

Domain knowledge can therefore be applied through a concise procedure (Autio et al., 2024).

- Build a minimal map: Collect the definitions, relationships, constraints, plausible values and key sources needed for the task.
- Identify invariants and edge cases: Establish which conditions cannot be violated and which exceptions might challenge a general formulation.
- Check orders of magnitude: Assess the plausibility of numbers, dates, percentages, durations and proportions.
- Check vocabulary: Check that technical terms are used consistently and with stable meanings.
- Reconstruct causal links: Make mechanisms, intermediate steps, conditions and contributing factors explicit.
- Identify critical claims: Select the statements that actually affect the meaning or validity of the work.
- Formulate targeted checks: Turn observed anomalies into specific questions, measurements, calculations or searches.
- Verify through independent references: Compare claims with sources, data, tests or expertise external to the model (Autio et al., 2024).

The procedure is not necessarily linear.

A source may change the initial model of the domain, while a contradiction may make it necessary to revise previously accepted terms or definitions.

Example
Suppose an *LLM* states:
“A person should drink at least four liters of water a day, because drinking more always improves hydration.”

The sentence is clear and sounds plausible, but minimal domain knowledge allows it to be analyzed more systematically.

- Build a minimal map: First, it is useful to know that water requirements vary according to age, physical activity, temperature, diet and individual conditions. It is also necessary to distinguish between water drunk directly and water consumed through food and other drinks (EFSA Panel on Dietetic Products, Nutrition, and Allergies [EFSA NDA Panel], 2010; Institute of Medicine, 2005).
- Identify invariants and edge cases: It is not plausible to assume that “more water” always means “better”. There are limits beyond which excessive intake can become problematic. The word “always” is therefore an immediate signal to check (EFSA NDA Panel, 2010; Institute of Medicine, 2005).
- Check orders of magnitude: Four liters a day might be reasonable under certain particular conditions, but cannot automatically be accepted as a universal value. The quantity therefore deserves specific verification (EFSA NDA Panel, 2010; Institute of Medicine, 2005).
- Check vocabulary: It is necessary to clarify what “hydration” means and whether the term is used consistently. Drinking more water does not automatically mean obtaining a proportional benefit.
- Reconstruct causal links: The sentence suggests the relationship “more water → better hydration”. However, the actual relationship depends on other factors, such as fluid losses, physical activity, temperature and the body’s ability to regulate fluid balance (EFSA NDA Panel, 2010; Institute of Medicine, 2005).
- Identify critical claims: There are two genuinely important statements: “at least four liters a day are needed” and “drinking more always improves hydration”.
- Formulate targeted checks: At this point, a generic search for information about water is unnecessary. It is better to ask: “Is there really a universal minimum amount of four liters?” and “Does continually increasing water intake always produce benefits?”
- Verify through independent references: The two claims can then be compared with health guidelines or other authoritative sources independent of the model that produced the response.

After verification, the initial sentence could be rephrased more cautiously:
“Water requirements vary between people and depend on several factors. Drinking an adequate amount is important for maintaining hydration, but consuming ever greater quantities does not necessarily produce additional benefits.” (EFSA NDA Panel, 2010; Institute of Medicine, 2005)

The example also shows why the procedure is not perfectly linear: during verification, we might discover, for example, that our initial idea of “daily requirements” was too simple and therefore need to revise the domain map before continuing the analysis.

It is important to remember that personal expertise is not a guarantee of correctness (Ericsson & Lehmann, 1996).
Even an expert can misremember, use an outdated definition, overlook a school of thought, apply an inappropriate criterion, be influenced by fluent text, overestimate their familiarity with a topic or accept as obvious something that is controversial in the field (Ericsson & Lehmann, 1996; Reber & Schwarz, 1999).

Domain knowledge should therefore be treated as a diagnostic and interpretive resource, not as an infallible authority (Autio et al., 2024).
Furthermore, the level of expertise required depends on the task.

- General knowledge may be enough to recognize an obviously implausible number.
- Evaluating a complex causal relationship may require a specialist (Ericsson & Lehmann, 1996).

When a task requires expertise one does not possess, the most responsible choice consists of the following actions (Autio et al., 2024):

- Consult authoritative sources.
- Ask a knowledgeable person for their opinion.
- Make uncertainty explicit.
- Avoid overly strong conclusions.
- Limit the practical use of the response (Autio et al., 2024).

Supplementary material

Exercises

Domain knowledge as a diagnostic tool
A generative system produces this explanation:
“In the HTTP protocol, every router that forwards a packet sends an ACK to the previous router to confirm that the data was received correctly.”
Analyze the response.

- Explain why the sentence may seem plausible to someone unfamiliar with networks.
- Indicate what minimum domain knowledge instead allows an anomaly to be identified.
- Distinguish the roles of:
    - HTTP.
    - Router.
    - IP.
    - TCP.
    - ACK.
- Explain why the problem does not necessarily lie in a single incorrect word, but in the relationship constructed between concepts belonging to different layers.
- Explain what is meant by a minimal domain map.
- Indicate which elements a good minimal map should contain:
    - Fundamental definitions.
    - Technical vocabulary.
    - Relationships between concepts.
    - Constraints and invariants.
    - Orders of magnitude.
    - Authoritative sources.
    - Edge cases.
- Explain why domain knowledge does not automatically demonstrate that the response is false.
- Formulate at least three verification questions more precise than a generic “Are you sure?”.
- Explain the sequence: Plausible output → domain knowledge → anomaly → targeted verification.
- Conclude by explaining why domain knowledge does not replace external sources, but guides verification.
-

Obvious error, subtle error or gray area?
Classify each of the following claims as:

- Obvious error.
- Subtle error.
- Gray area.

A: “A photograph measuring 1920 × 1080 pixels contains a total of 2,073 pixels.”
B: “Reducing JPEG quality decreases the image’s resolution.”

C: “A cool palette makes a scene melancholy.”
D: “A sequence shot increases immersion because it eliminates editing cuts.”

For each claim:

- Justify the classification.
- Identify what domain knowledge allows it to be evaluated.
- Indicate whether an invariant is violated.
- Identify any technical terms used imprecisely.
- Explain whether a quantitative sanity check is necessary.
- Identify any exceptions or missing conditions.
- Rewrite problematic claims more precisely.
- Explain why a subtle error can be more dangerous than an obvious error.
- Explain why a gray area should not automatically be treated as false.
- Conclude by explaining when domain knowledge is sufficient to flag a problem and when external verification becomes necessary instead.

Building a minimal map and using it to check an output

Choose one of the following domains:

- JPEG compression.
- Computer networks.
- Digital photography.
- UX Design.
- Audiovisual editing.

Before querying the model, build a minimal domain map containing at least:

- Five fundamental definitions or concepts.
- Five anchor words.
- Three important relationships between concepts.
- Three invariants.
- Two orders of magnitude or plausible values, where applicable.
- Two edge cases or exceptions.
- Three authoritative sources to use for any verification.

Then ask a generative system to explain a topic within the chosen domain.
Analyze the response.

- Highlight the main technical terms.
- Check that anchor words retain the same meaning throughout the response.
- Identify at least three claims that deserve attention.
- Check whether any invariants are violated.
- Perform a sanity check on the quantitative values present.
- Identify overly narrow or overly broad definitions.
- Look for exceptions the model may have ignored.
- Classify the anomalies as:
    - Obvious error.
    - Subtle error.
    - Gray area.
- Formulate a targeted checking question for each anomaly.
- Verify at least two claims through independent sources.

- Correct the response only after verification.
- Indicate which errors you probably would not have identified without the initial map.
- Conclude by assessing how effectively your initial domain knowledge served as a diagnostic filter.

Reconstructing a causal chain
A generative system states: “A sequence shot increases viewer immersion because it eliminates editing cuts.”

Do not immediately accept or reject the claim. Analyze it using domain knowledge.
- Identify the proposed cause.
- Identify the stated effect.
- Reconstruct at least three intermediate steps that could connect cause and effect.
- Construct a possible chain: Reduction/concealment of cuts → perceptual continuity → more stable construction of space and time → possible increase in engagement.
- Identify at least five conditions that could influence the effect:
    - Rhythm.
    - Camera movement.
    - Composition.
    - Acting.
    - Sound.
    - Genre.
    - Viewer expectations.
- Identify at least two cases in which a sequence shot could instead reduce engagement.
- Identify at least two cases in which editing with cuts could maintain or increase immersion.
- Explain why the initial formulation is causally incomplete, even if not necessarily false.
- Distinguish the consistency of the causal chain from evidence demonstrating it.
- Identify which steps in the chain require external evidence.
- Formulate at least three specific searches or verification questions.
- Rewrite the claim with a more appropriate degree of certainty.
- Finally, apply the procedure: Minimal map → invariants/edge cases → vocabulary → causal chain → critical claims → targeted checks → independent sources → rephrasing.
- Conclude by explaining why a highly convincing causal chain can still be false if it is not supported by adequate evidence.

Bibliography

Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). *Artificial intelligence risk management framework: Generative artificial intelligence profile* (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152. https://doi.org/10.1207/s15516709cog0502_2

Deering, S., & Hinden, R. (2017). *Internet Protocol, Version 6 (IPv6) specification* (RFC 8200). RFC Editor. https://doi.org/10.17487/RFC8200

Eddy, W. (Ed.). (2022). *Transmission Control Protocol (TCP)* (RFC 9293). RFC Editor. https://doi.org/10.17487/RFC9293

EFSA Panel on Dietetic Products, Nutrition, and Allergies. (2010). Scientific opinion on dietary reference values for water. *EFSA Journal, 8*(3), 1459. https://doi.org/10.2903/j.efsa.2010.1459

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Fielding, R., Nottingham, M., & Reschke, J. (2022). *HTTP semantics* (RFC 9110). RFC Editor. https://doi.org/10.17487/RFC9110

Hernán, M. A., & Robins, J. M. (2020). *Causal inference: What if*. Chapman & Hall/CRC. https://miguelhernan.org/whatifbook

Hill, A. B. (1965). The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine, 58*(5), 295–300. https://doi.org/10.1177/003591576505800503

Huang, J., Chen, X., Mishra, S., Zheng, H. S., Yu, A. W., Song, X., & Zhou, D. (2024). Large language models cannot self-correct reasoning yet. *International Conference on Learning Representations*. https://openreview.net/forum?id=IkmD3fKBPQ

Institute of Medicine. (2005). *Dietary reference intakes for water, potassium, sodium, chloride, and sulfate*. National Academies Press. https://doi.org/10.17226/10925

International Telecommunication Union. (1992). *Digital compression and coding of continuous-tone still images: Requirements and guidelines* (Recommendation T.81). https://www.itu.int/rec/T-REC-T.81/en

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248, 1–38. https://doi.org/10.1145/3571730

Jonauskaite, D., Abu-Akel, A., Dael, N., Oberfeld, D., Abdel-Khalek, A. M., Al-Rasheed, A. S., Antonietti, J.-P., Bogushevskaya, V., Chamseddine, A., Chkonia, E., Corona, V., Fonseca-Pedrero, E., Griber, Y. A., Grimshaw, G., Hasan, A. A., Havelka, J., Hirnstein, M., Karlsson, B. S. A., Laurent, E., … Mohr, C. (2020). Universal patterns in color-emotion associations are further shaped by linguistic and geographic proximity. *Psychological Science, 31*(10), 1245–1260. https://doi.org/10.1177/0956797620948810

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Further reading

- K. Anders Ericsson, Robert R. Hoffman, Aaron Kozbelt, and A. Mark Williams (eds.), The Cambridge Handbook of Expertise and Expert Performance, 2nd ed., Cambridge University Press, 2018 — Probably the strongest theoretical reference for this handout. It examines how expert knowledge differs from novice knowledge, how specialists represent problems, recognize meaningful patterns, make professional judgments, and develop domain-specific skills. It is particularly relevant to the handout's idea that expertise provides a structured conceptual map rather than merely a larger collection of isolated facts.
- Carl T. Bergstrom and Jevin D. West, Calling Bullshit: The Art of Skepticism in a Data-Driven World, Random House, 2020 — An excellent complement for the sections on orders of magnitude, numerical plausibility, misleading quantitative claims, causal interpretations, and sanity checks. The book develops practical habits for detecting claims that are formally impressive but conceptually or quantitatively weak. It closely matches the handout's recommendation to ask whether numbers, percentages, comparisons, and causal explanations are plausible before accepting them.
- Judea Pearl and Dana Mackenzie, The Book of Why: The New Science of Cause and Effect, Basic Books, 2018 — The strongest complement to the section on reconstructing causal chains. It explains why statistical association is different from causation and why causal claims require explicit models of how one condition could produce another. It is particularly relevant to the handout's insistence on identifying intermediate mechanisms, necessary conditions, alternative causes, and evidence supporting each step.
- Gary Klein, Sources of Power: How People Make Decisions, MIT Press, 1998 — Particularly useful for understanding expert recognition and professional judgment. Klein studies how experienced practitioners use domain-specific patterns and cues to identify important features of a situation rapidly. This fits the chapter's description of domain knowledge as a filter that can make an anomaly visible before the exact error has been formally demonstrated.
- Philip E. Tetlock and Dan Gardner, Superforecasting: The Art and Science of Prediction, Crown, 2015 — A strong complement for the final sections dealing with uncertainty, calibration, overconfidence, revision of beliefs, and the limits of expertise. It emphasizes decomposing problems, updating judgments when new evidence appears, and expressing conclusions with an appropriate degree of confidence rather than treating expertise as infallible.

Editorial notes

In producing these materials, I have used generative artificial intelligence tools to support the writing process, particularly to improve the form of the text, reorganize content, refine wording, and speed up certain editorial tasks.

As I work independently on the production of these materials, I try to automate any activities that can reasonably be automated, so that I can devote more time to research, design, and content development.

However, artificial intelligence does not determine the content of this work: the choice of topics, structure, ideas, interpretations, examples, and teaching approach are developed by me. AI is therefore used primarily as a tool to support production and formal revision, while authorship and responsibility for the design and content remain mine.

Unless otherwise indicated, this material is distributed under the Creative Commons Attribution–NonCommercial–ShareAlike 4.0 International license (CC BY-NC-SA 4.0).

You are therefore free to share, redistribute, adapt, and create derivative works based on this material, provided that appropriate credit is given, the material is not used for commercial purposes, and any modified or derivative versions are distributed under the same license.

