

AI vulnerability

Protect judgment, capability and responsibility in AI-assisted life.

DEFINITION

Vulnerability is contextual

A skill is not simply automated or safe. Vulnerability changes with the task, consequence, data, verification and the human capability that remains active.

TASK

What is delegated?

Drafting, searching, calculating, recommending, deciding or acting have different exposure.

WHY

What is at stake?

Convenience, learning, money, safety, rights and relationships require different oversight.

WHO

Who can verify?

Capability matters: a person cannot meaningfully supervise reasoning they do not understand.

OWN

Who is accountable?

Responsibility must remain visible even when the work is AI-assisted.

Score translation used in this book: The source field is named *aiVulnerabilityScore*, but the interface presents it as Future Relevance and “hard to replace” when high. To avoid reversing the meaning, any exposure number here is explicitly derived as **100 - FRI**.

CAPABILITY RETAINED THROUGH THE WORKFLOW



RESILIENCE + ADAPTABILITY + CRITICAL THINKING KEEP HUMAN JUDGMENT AVAILABLE

----- Risk path: delegate everything - accept fluency - lose explanation - become dependent -----

Design for transfer. A useful AI-assisted task should end with more independent understanding, not only a faster output.

FOUR PATHWAYS

How capability becomes vulnerable

The practical risk is not only job replacement. Capability can also weaken through routine delegation, over-trust and invisible responsibility.



VULNERABILITY CAN ENTER THROUGH MORE THAN ONE PATH
A visible tool may sit at the center, while exposure develops through delegation, dependency, over-trust or unclear ownership.

AUTOMATION EXPOSURE A task becomes easy to delegate, reducing the need to perform it manually. repeatable structured low-context	DESKILLING EXPOSURE A person stops practising the underlying reasoning and becomes less able to work independently. atrophy dependency shallow recall
OVER-TRUST EXPOSURE Fluent output is accepted without enough evidence, challenge or source checking. bias fabrication false confidence	ACCOUNTABILITY EXPOSURE It becomes unclear who made the choice, checked the result or owns the consequence. opacity handoff responsibility gap

RESEARCH PULSE • 2025 • R7

Immediate gains may not transfer

Four experiments with 3,562 participants found GenAI collaboration improved immediate performance, but gains did not persist on later solo tasks; intrinsic motivation also fell and boredom increased.

Wu et al. (2025), Scientific Reports • [Human-generative AI collaboration enhances task performance but undermines human's intrinsic motivation](#)

CONSEQUENCE MATRIX

Match oversight to consequence

The same tool behavior can be reasonable in one setting and unsafe in another. Verification effort should rise with consequence and irreversibility.

Low consequence · Easy to reverse Use AI freely for options, formatting and low-stakes experimentation. brainstorm outline rewrite	Higher consequence · Easy to reverse Use AI to compare options, then verify claims and test before committing. budget draft work plan public message
Low consequence · Hard to reverse Pause for consent, privacy and reputational effects even when the immediate stakes feel small. personal data image sharing identity claims	High consequence · Hard to reverse Use qualified human review and authoritative evidence. Do not delegate final judgment. safety rights health legal

Examples illustrate a decision method. They are not professional medical, legal or financial guidance.

DATA MATURITY

Read the dataset carefully

The current portfolio is rich in definitions and progressions, but several vulnerability fields still look like working defaults rather than completed assessments.

13 profiles in lowest FRI band Repeated working default; review before use	0 vulnerability tags populated across the 147 profiles	147 policies set to allowed no contextual variation yet
---	---	--

- 1

Use definitions

The strongest content is in skill definitions, behaviors, progressions and practice scenarios.
- 2

Use signals as prompts

The three interpretations can focus discussion, but should not drive a high-stakes conclusion alone.
- 3

Separate missing from low

A repeated value may indicate an incomplete review, not genuine low future relevance.
- 4

Add context

Future vulnerability work should assess tasks, consequences, verification and human oversight.

Editorial choice: This book does not publish a “most vulnerable skills” ranking because the current source does not support that level of confidence.

FOUNDATIONAL JUDGMENT

Protect the skills that verify reality

Some capabilities deserve deliberate practice because they let people check evidence, calculations, arguments and AI-generated claims.

SKILL AND ROLE		 HUMAN CORE	 FUTURE READINESS	 SAFEGUARD VALUE
Information Literacy Determines what evidence is needed, finds it, judges source quality and uses information ethically to answer the real question.	IN PRACTICE Identifies an evidence gap, compares independent sources and traces important claims to their origin. PROTECTS: misinformation, fabricated references and weak sourcing	 Human leads 4 / 5	 Strategic 4 / 5	 Strong safeguard 4 / 5
Critical Thinking Tests claims, assumptions and evidence before deciding what to believe or do.	IN PRACTICE Asks what supports the claim, seeks counterevidence and explains the final judgment. PROTECTS: automation bias, unsupported confidence and persuasive error	 Human leads 4 / 5	 Foundational 5 / 5	 Critical safeguard 5 / 5
Practical Numeracy Uses quantities, estimates, proportions and calculations to judge everyday claims and make sound decisions.	IN PRACTICE Estimates the expected range, checks units and independently reproduces the essential calculation. PROTECTS: numerical manipulation, calculation error and false precision	 Shared role 3 / 5	 Strategic 4 / 5	 Strong safeguard 4 / 5
Logical Reasoning Evaluates whether conclusions follow from their premises by identifying contradictions, invalid steps and common fallacies.	IN PRACTICE States the premises, tests the inference and looks for a counterexample that would break it. PROTECTS: plausible nonsense, internal contradiction and invalid inference	 Human leads 4 / 5	 Strategic 4 / 5	 Strong safeguard 4 / 5

Capability test: If the tool were unavailable, could you still explain the evidence, reproduce the essential calculation or detect a contradiction?

DECISION GATES

Keep a human in the loop

Human oversight becomes real through explicit gates, not through a generic promise that a person is involved.

1

Intent gate

Is the goal legitimate, clear and aligned with the people affected?

2

Data gate

Can this information be shared? Is privacy, consent or confidentiality involved?

3

Evidence gate

Which claims, sources, calculations or assumptions require independent checking?

4

Consequence gate

What could happen if the output is wrong, biased, leaked or misunderstood?

5

Accountability gate

Who makes the final decision, signs off and can explain the reasoning?

6

Learning gate

What should the person still be able to do without the tool after this task?

RESEARCH PULSE · 2025 · R8

Explanations alone do not prevent automation bias

A review of 35 studies found that explainability can increase acceptance without reliably improving accuracy. Active engagement and independent verification are stronger safeguards.

Romeo & Conti (2025), AI & SOCIETY · [Exploring automation bias in human-AI collaboration: a review and implications for explainable AI](#)

USE MODES

Choose mode by consequence

Move from open exploration to strict human control as uncertainty, sensitivity and consequence increase.

MODE	APPROPRIATE USE	HUMAN RESPONSIBILITY
USE	Explore and vary Ideas, outlines, formatting and reversible low-stakes work.	Set intent and select.
ASSIST	Support skilled work Drafting, comparison, coding, analysis and planning.	Understand, test and adapt.
VERIFY	Consequential claims Evidence, recommendations, calculations and public communication.	Check independently and document.
OWN	Final human judgment Rights, safety, health, legal status, consent and accountability.	Qualified person decides.

Context beats blanket policy: The same AI system may move between modes inside one workflow.

RESEARCH PULSE · 2025 · R9

Deskilling needs longitudinal monitoring

A 2025 review in medicine identified risks to judgment, examination and professional autonomy from AI-supported work. Its authors call for longitudinal monitoring and explicit safeguards against skill erosion.

Natali et al. (2025), Artificial Intelligence Review · [AI-induced Deskilling in Medicine: A Mixed-Method Review and Research Agenda for Healthcare and Beyond](#)

Contextual policy model created for this minibook. The source database currently marks all profiles “allowed,” without consequence-specific rules.

ANTI-DESKILLING

Keep the capability alive

A good AI workflow should leave the person more able to explain, verify and act - not merely faster.

KEEP INDEPENDENT UNDERSTANDING AVAILABLE

AI can illuminate the work without becoming the only source of direction. The person must still be able to explain, verify and choose.

- 1

Attempt first

Write a hypothesis, outline, calculation or plan before asking for assistance.
- 2

Request friction

Ask for critique, counterarguments, missing evidence and questions - not only an answer.
- 3

Test the output

Use another source, a known example, a calculator, a peer or a small experiment.
- 4

Explain it back

Reconstruct the logic and key facts from memory in your own words.
- 5

Make the call

Choose what to use, change or reject. State the reason.
- 6

Log the learning

Record one thing gained, one uncertainty and one capability to practise next.

Minimum standard: If you cannot explain why the result is credible and appropriate, you are not yet ready to rely on it.

SCENARIO PRACTICE

Make the stronger pattern concrete

Vulnerability becomes easier to manage when we compare observable behavior in the same situation.

CONTEXT	CAPABILITY SHRINKS	CAPABILITY GROWS
01 Learning	Vulnerable pattern Submit an AI summary you cannot explain.	Stronger pattern Read first, create questions, compare the AI summary and explain the concept without it.
02 Work	Vulnerable pattern Forward a polished recommendation without checking assumptions.	Stronger pattern Identify decision criteria, verify key claims and record the accountable sign-off.
03 Personal life	Vulnerable pattern Treat a confident answer as professional advice.	Stronger pattern Use AI to prepare questions, then consult authoritative sources or a qualified person when stakes are high.

Change the workflow, not only the attitude. Add a first attempt, a verification step and an accountable decision.

PERSONAL AUDIT

Where is your capability exposed?

Complete the audit for one recurring AI-assisted task. The aim is to redesign the workflow, not to stop useful technology.

Audit statement	No	Sometimes	Yes
☐ I can complete the essential task without AI when necessary.	No	Sometimes	Yes
☐ I understand what information the system received.	No	Sometimes	Yes
☐ I know which claims or calculations require verification.	No	Sometimes	Yes
☐ I can explain the reasoning in my own words.	No	Sometimes	Yes
☐ I have considered bias, missing context and affected people.	No	Sometimes	Yes
☐ The level of oversight matches the consequence.	No	Sometimes	Yes
☐ The final accountable person is named.	No	Sometimes	Yes
☐ I notice signs of dependency or loss of confidence.	No	Sometimes	Yes
☐ I deliberately practise the underlying human skill.	No	Sometimes	Yes
☐ I review failures and update the workflow.	No	Sometimes	Yes

Stop

One delegation that hides risk or weakens learning.

Start

One gate, verification or explanation behavior.

Continue

One use where AI clearly expands capability.



The goal is not to avoid AI. It is to remain capable with it.

Keep enough skill to challenge the output, enough judgment to match oversight to consequence, and enough agency to own the final decision.

About This Material

WHAT THIS IS

This material is part of the Skills Pathfinder learning ecosystem developed by GrowthMindsetBuilder. It presents one or more human skills in a concise, practical format for learning, reflection and training.

It may be used as a standalone learning resource, but it is not a complete course. The broader learning experience—including facilitated activities, practice, feedback and progression—lives within the programmes and client projects from which this material was developed.

RESEARCH FOUNDATION

Skills Pathfinder profiles and frameworks synthesize extensive research, international models and established good practices. They bring together multiple perspectives to make complex ideas accessible and useful in real-life learning and development.

No individual model or study provides a complete answer. Skills Pathfinder should therefore be understood as an evidence-informed framework for reflection and growth—not as a universal or final definition of human capability.

RESEARCH NOTES AND SOURCES

Where used, research callouts marked [R1], [R2] and so on provide short summaries of published studies. Each callout identifies the authors and publication so that you can find and examine the original research.

Research on human skills, AI capabilities and future readiness is developing quickly. Findings may differ across people, tasks, technologies and settings—and conclusions that appear settled today may change as new evidence emerges.

Treat each study as one contribution to the evidence rather than a final verdict. Compare perspectives, explore the original sources and form your own view. Where a source is cited, the original publication—not this summary—should be treated as authoritative.

CONTEXTUAL USE

Every skill profile is contextual. Language, terminology, emphasis and interpretation should be adapted to the cultural, educational, professional and organizational setting in which the material is used.

Local knowledge—and the experience of the people using the framework—is essential for making the material relevant, inclusive and effective.

EDUCATIONAL USE AND DISCLAIMER

This material is provided for educational and developmental purposes. It is not medical, legal, financial, psychological or other professional advice.

Reflection tools and self-assessments are intended to support learning and conversation. They are not clinical, psychological or psychometric diagnostic instruments and should not be used as the sole basis for employment, education or other high-stakes decisions.

Unless stated otherwise, case stories are composites created for learning use and do not describe identifiable individuals.

Frameworks and concepts credited to other authors remain their intellectual work. The practice, reflection and progression language surrounding them represents GrowthMindsetBuilder's interpretation and application.

Markku Pelkonen

GrowthMindsetBuilder · Growth Resilience Oy

markku.pelkonen@growthmindsetbuilder.com

+358 45 318 7300

growthmindsetbuilder.com · skillspathfinder.com

© Growth Resilience Oy. Shared for learning and training use. If you quote or reproduce an extract, please retain the attribution and relevant source notes.