Co-Designing a Checklist to Assess the Risk of AI-Induced Skill Erosion in Professional Knowledge Work

Angrisani Tommaso
Politecnico di Torino
Torino, Italy

Carpenzano Giacomo
Politecnico di Torino
Torino, Italy

Cesaria Francesco Luigi Karol
Politecnico di Torino
Torino, Italy

Del Debbio Luca
Politecnico di Torino
Torino, Italy

Gallea Luca
Politecnico di Torino
Torino, Italy

Preti Andrea
Politecnico di Torino
Torino, Italy

Protasio Brend Margarette
Politecnico di Torino
Torino, Italy

Saladini Lorenzo
Politecnico di Torino
Torino, Italy

Scimone Tiziano
Politecnico di Torino
Torino, Italy

Sdiri Mohamed
Politecnico di Torino
Torino, Italy

Abstract

The integration of generative artificial intelligence (AI) into professional workflows is quietly reshaping not just how work gets done, but whether workers retain the ability to do it without AI at all. Existing responsible AI frameworks address fairness and transparency at the system level but offer no practical instrument for detecting when AI adoption is eroding the human expertise. This paper presents a co-designed checklist for assessing AI-induced deskilling risk in professional knowledge work. Grounded in our foundational scoping review (which analyzed 36 materials), the tool was refined through a multi-phase iterative process involving 11 engineering students and 10 expert reviewers for structural co-design and 25 professional practitioners who field-tested the evolving checklist across its three iterations. Mapping feasibility across three risk pillars (individual cognitive mechanisms, professional development, and organizational factors) revealed a clear pattern: individual verification habits were easy to adopt, but organizational items hit a wall of deadline pressure, misaligned incentives, and absent governance. Workers know the risk. Their organizations are not built to let them act on it. The final instrument, V3.0, uses scaled frequency questions and open text fields that require respondents to explain their reasoning, creating a genuine reflection rather than box-ticking. A scoring framework then maps each response to an individual autonomy profile and an organizational archetype, giving teams a concrete picture of where deskilling risk is coming from and what specifically needs to change.

Keywords

AI Task Delegation, Skill Erosion, Co-Design, Deskilling, Professional Expertise, Human Judgment, Checklist.

1 Introduction

While a growing body of research has documented the mechanisms of AI-induced deskilling, from cognitive offloading to the erosion of tacit knowledge, the vast majority of this work remains diagnostic rather than actionable. Practitioners and managers are left with a clear understanding of the problem but no structured way to assess whether their own workflows are actively eroding expertise, nor any practical guidance on what to change. Existing responsible AI frameworks address fairness, transparency, and accountability at the system level, but they offer no instrument for detecting when routine AI delegation is silently undermining the very human capabilities that organizations depend on for innovation, error recovery, and strategic judgment. This gap between abstract risk awareness and concrete daily practice is not merely academic: without a structured reflection tool, teams default to efficiency-driven behaviours, and deskilling accumulates unnoticed until a system failure or an unassisted task reveals the loss. To bridge this gap, we developed a co-designed checklist that translates the three pillars of skill erosion identified in our scoping review (individual cognitive mechanisms, professional development risks, and organizational factors) into a practical, self-administered instrument. The checklist does not assume that deskilling can be eliminated, but it provides a replicable way for knowledge workers and their organizations to see the risk, locate its source, and act before expertise is irreversibly compromised.

2 Methods

We followed an iterative co-design process grounded in the methodology described by Madaio et al. [2] to develop and refine our checklist. This section describes the data collection and analysis procedures.

2.1 Participants

To ensure a robust and comprehensive co-design process, we engaged three distinct participant cohorts across different phases of the project:

- **Initial Peer Review:** During the early synchronous stage, 11 engineering students from other research groups evaluated the baseline constructs to eliminate academic jargon.

- **Expert Co-Design Panel (N=10):** A targeted panel of 10 professional practitioners acted as structural co-designers. Four experts evaluated the transition from Draft V1.0 to V2.0 (focusing on technical feasibility), and an additional four guided the refinement from V2.0 to the final V3.0 (focusing on usability and scoring calibration).
- **End-User Testers (N=25):** To gather empirical data and validate the tool in real-world scenarios, the evolving checklist was deployed to 25 independent knowledge workers. Specifically, 7 participants tested V1.0, 10 tested V2.0, and 8 evaluated the final V3.0. To ensure broad generalizability, this cohort was drawn from highly diverse professional domains, including healthcare, law, education, human resources, software engineering, and digital marketing.

2.2 Materials

The co-design protocol utilized a progression of physical and digital materials structured to facilitate iterative refinement. In the synchronous phase (Phase 1), the primary materials consisted of three physical A3 paper sheets, each dedicated to one of the risk macro-areas derived from the scoping review (individual cognitive mechanisms, professional development, and organizational factors). These sheets served as tangible shared spaces for hosting peer feedback via physical sticky notes.

In the asynchronous phase (Phase 2), the refined toolkit was transitioned into digital formats: a Google Doc file was used for collaborative text editing and linguistic harmonization. This digital survey was structurally operationalized using a binary multiple-choice format (representing green and red feasibility indicators) paired with mandatory open-ended text fields to collect qualitative co-design suggestions from external experts.

2.3 Data Analysis

We adopted an inductive thematic analysis approach, built upon the methodological steps described by Madaio et al. [2], to systematically process the empirical feedback gathered across both co-design phases. The evaluation framework combined both quantitative metrics from the external validation loop and qualitative insights linked to our three theoretical risk pillars.

Accordingly the participants' recommendations were systematically categorized into three actionable design interventions:

(1) **Additions**, integrating missing socio-cognitive or structural risks highlighted by practitioners;

(2) **Removals**, eliminating checkpoints deemed redundant or impractical within real-world data science workflows;

(3) **Language changes**, rephrasing items saturated with academic jargon into clear, professional corporate vernacular.

This analytical sequence helped us track everything and be open, passing the test of being clear. It directly guided us to improve our Checklist from the initial version to the final version 3.0.

2.4 Overview of the Co-Design Process

The co-design process followed an iterative, multi-stage architecture designed to translate the theoretical insights from our scoping review into an operational tool, combining synchronous, in-person co-design sessions (peer-to-peer co-design) with a subsequent asynchronous digital scrutiny phase involving external stakeholders.

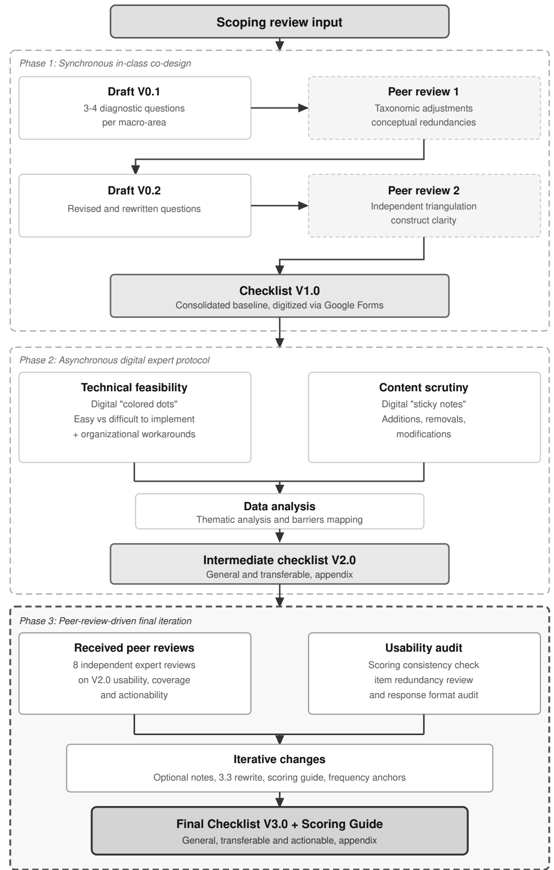


Figure 1: Summarizes our co-design procedure, adapted from the AI Impact Assessment Report Template co-design framework [2]

2.4.1 *Phase 1: Synchronous In-Class Co-Design (Iterative Peer Review).*

The conceptual core of the checklist was initially divided into three distinct physical A3 paper sheets, mapping directly to the risk macro-areas identified in our review: (1) Individual Cognitive Mechanisms, (2) Professional Development, and (3) Structural & Organizational Factors. For each macro-area, the team drafted 3 to 4 initial diagnostic questions (Version 0.1). The refinement protocol consisted of two consecutive rounds of critical evaluation by engineering peers from other research groups within the university course, acting as proxy evaluators.

- **First Review Round ($V0.1 \rightarrow V0.2$):** Peer reviewers analyzed the A3 sheets and applied physical yellow sticky notes to suggest taxonomic adjustments, highlight conceptual redundancies, or propose missing items. Based on this feedback, the team revised and rewrote the questions onto a new set of A3 sheets.
- **Second Review Round ($V0.2 \rightarrow V1.0$):** The amended version underwent a second independent triangulation with a different

group of peers. This round focused primarily on construct clarity and the elimination of academic jargon. This phase resulted in our first consolidated baseline checklist (Version 1.0), which was transcribed into a master Google Doc.

2.4.2 Phase 2: Asynchronous Digital Protocol for Technical and Compliance Experts. To validate the tool's real-world utility, Version 1.0 was digitized via Google Forms and deployed to a targeted sample of external corporate practitioners. To prevent the phenomenon of "compliance theatre" (where practitioners passively check boxes without deep reflection), the form was structured as an interactive feedback infrastructure:

(1) **Technical Feasibility Assessment (Digital "Colored Dots"):** For each checklist item, experts evaluated the friction of implementing that specific check within their actual daily work routines using a binary multiple-choice format: (Green dots) Easy to implement / Already in practice and (Red dots) Impossible, irrelevant or important but difficult to implement (e.g., due to tight deadlines, manager pressure, or lack of time). In alignment with institutional feedback guidelines, a dedicated section was provided for Red selections, allowing the expert to specify organizational workarounds or managerial pivots required to make the human oversight viable.

(2) **Content Scrutiny (Digital "Sticky Notes"):** At the end of each of the three macro-sections, mandatory open-ended text boxes acted as digital sticky notes. Participants were instructed to act as co-designers, providing targeted feedback for Addition (missing socio-cognitive risks), Removal (corporate-unfriendly or redundant checkpoints), or Modification.

3 Findings and Inductive Analysis Results

The execution of the three-phase co-design protocol yielded a robust dataset of both quantitative operational metrics and qualitative structural evaluations. This section details how this empirical evidence guided the structural evolution of the checklist, bridging the gap between its initial theoretical design (Version 1.0) and the final, practitioner-validated instrument (Version 3.0).

3.1 Quantitative Feasibility Mapping (Friction Scale)

During Phase 2, expert stakeholders and industry practitioners were asked to evaluate the operational feasibility of each checklist item by applying it to simulated professional workflows through storyboarding. Using the Digital Colored Dots method, participants categorized items based on the "friction" their deployment would introduce into their daily professional routines. The quantitative mapping revealed a clear polarization between two distinct categories of items:

- **Low-Friction Items (Green Dots):** Questions focused on individual cognitive habits, such as critical verification (Item 1.1), received highly positive ratings. Practitioners felt that verifying code snippets or calculated statistics is already deeply embedded in the ethics of data work. Formalizing these checks was viewed as a helpful reinforcement of existing standards rather than a workflow bottleneck.
- **High-Friction Items (Red Dots):** Conversely, structural and organizational factors exhibited a high density of friction indicators. Specifically, Item 3.1 (Time Constraints) and Item 2.3 (No-AI Exercises) were flagged as highly challenging to execute. For example, enforcing regular "No-AI" exercises to preserve raw skills (Item 2.3) was flagged as theoretically ideal but practically very difficult. As one participant bluntly noted, "I try yes, but not under deadline." Participants highlighted a systemic corporate conflict: managerial pressure for sheer output volume often overrules the time required to critically audit AI outputs.

3.2 Qualitative Checklist Evolution

Inductive coding of qualitative feedback highlighted three primary categories of design intervention.

Additions were driven by industry experts who emphasized capturing passive automated compliance rather than just active delegation; this prompted the refinement of Item 1.2 to measure exposure to superficially correct AI outputs and the addition of Item 3.5 to address data privacy and intellectual property governance.

Removals eliminated technical anchors that heavily verticalized the initial draft (V1.0) toward specific languages (e.g., direct mentions of SQL syntax or R dataframes), thereby elevating the framework into a role-agnostic instrument scalable across broader knowledge-work domains.

Finally, **Language Changes** translated abstract academic jargon into an operational corporate vernacular: the theoretical concept of "Automation Bias" was rephrased as "passively trusting an AI output" (Item 1.2), while "Never-skilling" and "loss of Tacit Knowledge" were operationalized into the empirical observation of junior colleagues losing foundational skills like reading raw logs or debugging simple scripts (Item 2.2). While the transition to V2.0 resolved initial structural friction, the second panel of expert reviewers highlighted new usability and analytical limitations. First, reviewers noted visual fatigue caused by inconsistent multiple-choice layouts; consequently, V3.0 standardized the direction of all frequency scales (e.g., aligning the lowest-risk scenarios consistently on the same side) to prevent data-entry mistakes. Second, redundancies within organizational factors (such as overlapping questions on speed vs. quality) were merged to streamline the tool. Finally, participants noted that the checklist lacked a clear operational output. To address this, the transition to V3.0 introduced the dual-axis Scoring Guide, translating qualitative answers into quantifiable Individual Autonomy and Organizational Environment archetypes. Furthermore, addressing feedback regarding "survey fatigue" under tight deadlines, mandatory justification fields were modified to "Notes (Optional)" in V3.0, preserving the opportunity for deep reflection without imposing prohibitive time constraints. The complete evolution, mapping these qualitative and structural transformations from Draft V1.0 to the final V3.0, is comprehensively documented in the Visual Changelog (Table 1)

Table 1: Visual Changelog: Iteration from V1.0 to the final V3.0 based on Stakeholder Feedback

Design Goal	Previous Draft & Stakeholder Feedback	Action	Optimized Item / Framework Feature
Disrupting ComplianceTheatre	*Draft V1.0:* Relied heavily on binary "Yes/No" checkpoints (e.g., "Do you have regular 'No-AI' exercises?"). **Feedback (V1):** "A simple 'Yes' doesn't capture the reality of our workflow. It's too easy to just tick the box without reflecting."	**Format Shift**	**Item 2.3 (V2.0):** Transitioned binary checkpoints into 5-point behavioral frequency scales (Never → Always) to enforce genuine reflection.
De-jargonizing Cognitive Risks	*Draft V1.0:* Framed questions around theoretical academic concepts like "Automation Bias" and "Cognitive Offloading". **Feedback (V1):** Corporate practitioners found the wording abstract and disconnected from their daily routines.	**Language Change**	**Item 1.2 (V2.0):** Rephrased abstract risks into relatable, empirical scenarios (e.g., "How often do you catch yourself passively trusting an AI output because it looks superficially correct?").
Expanding Structural Scope	*Draft V2.0:* Structural factors focused on workload and internal incentives, missing external compliance pressures. **Feedback (V2):** Legal/Compliance stakeholders highlighted that unsupervised AI reliance often stems from absent IP and privacy guidelines.	**Structural Addition**	**Item 3.5 (V3.0):** Added a specific checkpoint to evaluate the presence of "comprehensive governance policies" regarding intellectual property and data privacy.
Actionability & Usability	*Draft V2.0:* The checklist provided rich qualitative data but no diagnostic outcome. Furthermore, mandatory "Notes" caused survey fatigue. **Feedback (V2):** "We need a way to interpret the final result quickly. Also, skip mandatory typing under tight deadlines."	**Diagnostic Upgrade**	**V3.0 Final:** Changed text fields to "Notes (Optional)" to reduce friction. Implemented the dual-axis *Scoring Guide* mapping users to specific Deskilling Archetypes.

4 Discussion

4.1 The Organisational Nature of Deskilling Risk

The feasibility mapping from Phase 2 revealed a clear pattern: individual-level items—such as critical verification of AI outputs—were rated as easy to implement, with practitioners noting that such checks align with existing professional norms. In contrast, organisational items consistently hit a wall of structural friction. Deadline pressure (Item 3.1), misaligned incentives (Item 3.3), and the absence of formal "Plan B" procedures (Item 3.2) were flagged as impossible or impractical under real-world corporate workflows. As one participant bluntly noted, "I try yes, but not under deadline." This finding challenges the common assumption that deskilling is primarily an individual cognitive failing. Instead, our data indicate that even highly motivated practitioners are systematically forced into passive AI acceptance when their organisational environment rewards speed over verification and provides no structural buffer for critical review. Overcoming this erosion trap therefore requires more than individual vigilance; it demands that organisations redesign incentive systems, allocate protected time for output auditing, and embed the checklist into existing governance touchpoints such as procurement pipelines or quarterly sprint retrospectives. Without these structural changes, any deskilling mitigation strategy will remain aspirational.

4.2 From Compliance Theatre to Genuine Reflection

Early versions of the checklist (V1.0 and V2.0) used binary yes/no formats, which risked reducing the instrument to a superficial box-ticking exercise. Practitioners could quickly affirm that they "verify AI outputs" without revealing the frequency, depth, or conditions of that verification. To disrupt this performative use, the final instrument (V3.0) replaces binary questions with scaled frequency responses (e.g., Never / Rarely / Sometimes / Often / Always). Furthermore, while earlier iterations tested mandatory justification fields, V3.0 transitions to "Notes (Optional)" to balance the collection of qualitative context with the reality of survey fatigue under tight corporate deadlines. Even without mandatory writing, the frequency scales introduce deliberate cognitive friction: respondents cannot simply check a default "Yes" but must actively weigh their actual behavioral patterns. Finally, the scoring framework transforms individual responses into two actionable profiles: an Individual Autonomy Score (mapping to archetypes such as "Vulnerable Synergist") and an Organizational Environment Score (e.g., "Erosion Trap" or "Skill-Safe Ecosystem"). This dual feedback mechanism forces teams to confront not only their own habits but also the systemic conditions that shape them, turning the checklist from a passive audit into a lever for genuine organizational reflection.

4.3 Limitations and Future Work

Several limitations must be acknowledged. First, while our field-testing successfully engaged a diverse cohort of 25 practitioners, spanning fields from radiology to legal assistance and creative design, the sample size within each specific profession remains inherently small. This broad variance demonstrates the tool's generalizability but prevents us from drawing definitive conclusions about how deskilling risks might structurally differ across specific industries. Second, while the end-user base was highly diverse, the expert co-design panel (N=10) primarily consisted of operational practitioners. Future iterations should explicitly incorporate higher-level management and compliance staff into the structural review process, as organizational items are precisely where managerial buy-in is most critical.

5 The Finalized Checklist Framework (V3.0)

The iterative co-design process culminated in the consolidation of a structured evaluation tool organized into four macro-logical areas, transitioning from a rigid binary architecture (Yes/No) to nuanced behavioral options and frequency Likert scales paired with open-text qualitative fields.

The framework maps structural control variables within the Participant Profile (professional role, seniority, and baseline AI usage frequency) to contextualize subsequent evaluations. Short-term micro-risks are addressed under **Individual Cognitive Mechanisms**, analyzing immediate behavioral interactions such as verification patterns, automation complacency, and problem-solving approaches (manual error diagnosis versus passive regeneration).

Long-term expertise decay is evaluated under **Professional Development Risks**, capturing empirical signs of never-skilling, logic retention, and baseline competency erosion.

Finally, **Structural Factors** investigate organizational capacity, workload friction, incentive alignment, and formal data governance policies. V3.0 also introduces a scoring guide (see Appendix) that maps responses to individual autonomy profiles and organisational archetypes, enabling teams to diagnose not just whether deskilling risk exists, but where it originates and what specifically needs to change.

6 Worked Use Case

A mid-level data analyst at a major streaming platform is tasked with a technically demanding routine: developing a daily Python pipeline to process listening session data, calculate rolling 7-day engagement metrics, and integrate results into a production dashboard. Tempted to delegate the entire script via generative AI, the analyst instead employs the co-designed checklist as a reflective gate before proceeding.

The evaluation starts with **Item 1.1**: superficial code reviews won't catch subtle window-function logic; manual cross-checking against known data is required. **Item 1.2** brings back a memory: it has happened in the past where an AI-generated piece of code looked correct but introduced hidden errors throughout the pipeline. Furthermore, **Item 1.4** reveals a behavioural shift—the analyst has started asking the AI to "just try again" rather than debugging manually, a pattern that has visibly eroded their ability to fix similar issues alone.

Moving to Professional Development, **Item 2.4** shows they can no longer draft a basic partition window from memory without looking up the syntax. **Item 2.3** exposes a structural gap: there are no formal "No-AI" exercises and no mandated manual practice. Finally, **Item 3.1** surfaces a systemic conflict: performance metrics reward dashboard uptime and delivery speed, not rigorous auditing of AI outputs. There is no incentive for catching errors, only for shipping fast.

The analyst then computes the scoring framework. The **Individual Autonomy Score** lands at 11 points, which is within the *Vulnerable Synergist* profile (11–21 points): task comprehension is intact, but verification patterns are sporadic. The **Organizational Environment Score** totals −3 points, placing the environment in the *Erosion Trap* (−5 to −2): the workplace culture penalises reflection, forcing even motivated practitioners into cognitive offloading just to survive deadlines. To visualize this assessment, the analyst's final scores are mapped onto the Skill-Safe Framework Matrix (Figure 2). The red marker at coordinates (-3, 11) visually confirms the precarious position of the use case: while the analyst retains enough foundational knowledge to avoid complete AI dependence (Profile A), the hostile organizational environment exerts a continuous downward pressure, increasing the risk of long-term deskilling. Furthermore, by overlaying the empirical data from the eight end-user testers who compiled the V3.0 checklist (P1–P8) onto the same matrix, the broader diagnostic utility of the tool becomes evident. The distribution reveals a concerning industry trend: half of the sampled professionals cluster in the "Vulnerable Synergist" tier, frequently constrained within the Erosion Trap or the Friction Zone. This visualization proves that the checklist not only evaluates individual tasks but also highlights how AI-induced deskilling is often a systemic failure rather than a purely individual

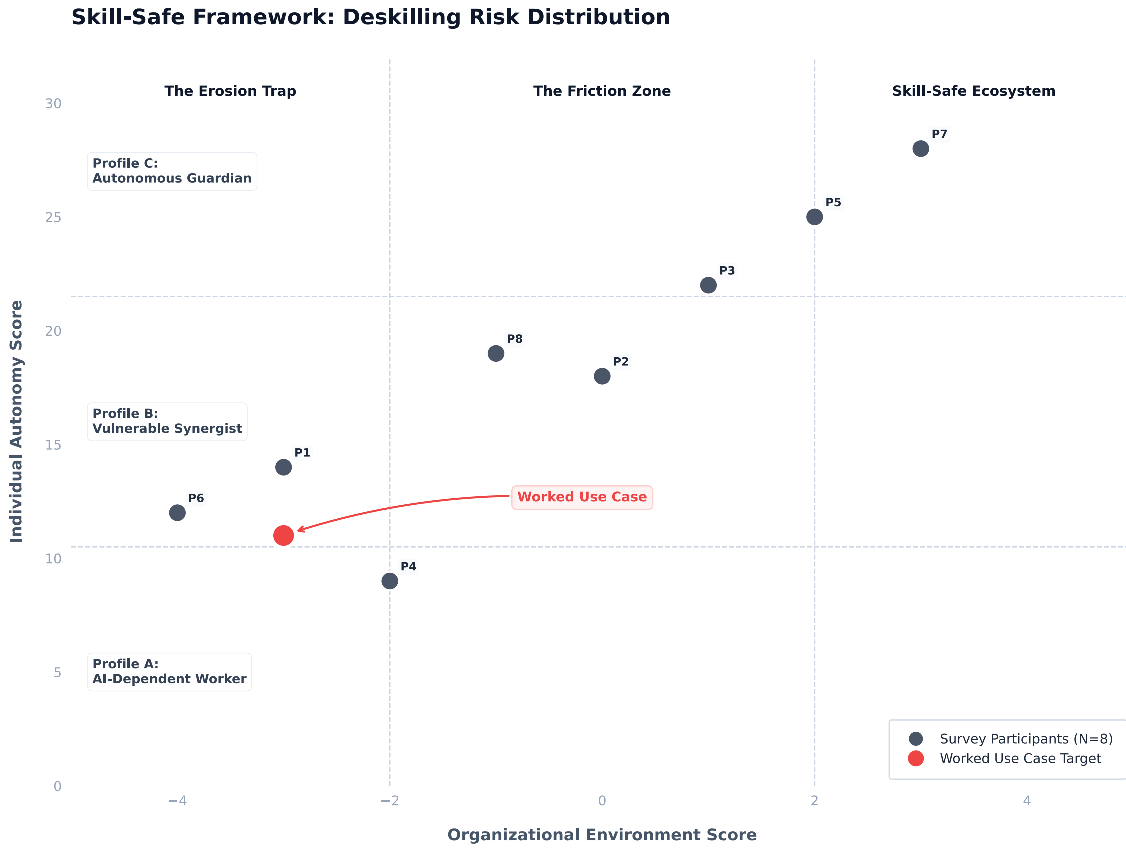


Figure 2: The Skill-Safe Framework Matrix. The chart plots the distribution of deskilling risk based on Individual Autonomy and Organizational Environment scores. The red marker isolates the Worked Use Case within the Erosion Trap. The dark gray markers (P1–P8) represent the empirical distribution of the eight co-design survey participants, demonstrating the framework's capability to visually diagnose systemic vulnerabilities across a broader workforce.

shortcoming, providing managers with a clear map for targeted organizational interventions.

7 Conclusions

Our findings reveal a clear pattern: individual verification habits are easy to adopt, but organizational factors hit a wall of deadline pressure, misaligned incentives, and absent governance. Workers know the risk. Their organizations are not built to let them act on it.

To break this pattern, the V3.0 Checklist introduces a scoring guide that acts on both the workers and the organizations, forcing genuine reflection. Its dual scoring framework, in fact, separates individual autonomy from organizational context, giving teams a shared language to diagnose whether deskilling stems from personal habits, workplace pressure, or both.

We recommend embedding the checklist into existing governance touchpoints—procurement, retrospectives, skills audits—rather than deploying it as daily micro-control. Mitigating deskilling is not an individual duty; it is a structural choice. The Skill-Safe Framework makes that choice visible, traceable, and actionable.

References

[1] Angrisani, T., Carpenzano, G., Cesaria, F. L. K., Del Debbio, L., Gallea, L., Mossetto, F., Preti A., Protasio Brend, M., Saladini, L., Scimone, T., Sdiri, M. 2026. The Risks of Skill Erosion and Judgment Loss Under AI Task Delegation: A Scoping Review on Professional Development. *Politecnico di Torino.*

[2] Madaio, M. A., Stark, L., Wortman Vaughan, J., Wallach, H. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp.1-14). ACM. https://doi.org/10.1145/3313831.3376445

[3] Constantinides, M., Bogucka, E., Quercia, D., Kallio, S., Tahaei, M. 2024. RAI Guidelines: Method for Generating Responsible AI Guidelines Grounded in Regulations and Usable by (Non-) Technical Roles. In *Proceedings of the ACM on Human-Computer Interaction (CSCW).* ACM. https://doi.org/10.1145/3686927

[4] Bogucka, E., Constantinides, M., Šćepanović, S., Quercia, D. 2024. Co-designing an AI Impact Assessment Report Template with AI Practitioners and AI Compliance Experts. *In Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society (AIES '24).* https://hdl.handle.net/11583/2996100

Appendix

AI Usage Declaration

Tools used: ChatGPT (OpenAI), Gemini Pro (Google), Claude (Anthropic), DeepSeek (DeepSeek).

How they were used: During the co-design process, LLMs were used to paraphrase early versions of checklist items, to identify ambiguous phrasing, and to suggest alternative wordings that reduced academic jargon. For the final report, LLMs helped with LaTeX formatting, minor copy-editing, and the structural organisation of the appendix.

Verification performed: All AI-generated suggestions—whether for search terms, paraphrased items, or editorial changes—were reviewed, fact-checked, and either accepted or rejected by at least two team members. No AI-generated content was included without explicit human validation. Every source cited in the scoping review was located, read in full, and manually extracted by a team member. LLMs were not used to generate core arguments, interpret findings, or replace stakeholder feedback. The final checklist (V3.0) and its scoring framework are entirely human-designed, with AI used only as a linguistic and formatting aid.

Statement of accountability: The authors assume full responsibility for the accuracy, originality, and integrity of the work.

Group: Me, Myself, AI

- **Angrisani Tommaso**: s340221@studenti.polito.it
- **Carpenzano Giacomo**: s342042@studenti.polito.it
- **Cesaria Francesco Luigi Karol**: s341586@studenti.polito.it
- **Del Debbio Luca**: s342568@studenti.polito.it
- **Gallea Luca**: s341964@studenti.polito.it
- **Preti Andrea**: s297321@studenti.polito.it
- **Protasio Brend Margarette**: s339680@studenti.polito.it
- **Saladini Lorenzo**: s322340@studenti.polito.it
- **Scimone Tiziano**: s337525@studenti.polito.it
- **Sdiri Mohamed**: s341348@studenti.polito.it

The Checklist Framework

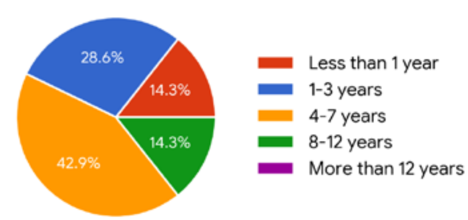
All three versions of the Checklist Framework are comprehensively documented below:

Checklist V1.0

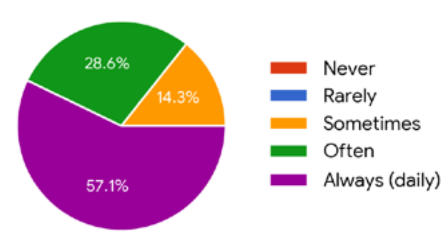
Deployment Metrics: *This initial draft was compiled and field-tested by* ***7 End-User Testers****. It was subsequently reviewed by a panel of* ***5 Expert Co-designers****, whose structural feedback drove the iteration to Version 2.0.*

0.1 What is your primary professional role? ____________

0.2 How many years have you been working in your field?
- ○ Less than 1 year
- ○ 1–3 years
- ○ 4–7 years
- ○ 8–12 years
- ○ More than 12 years

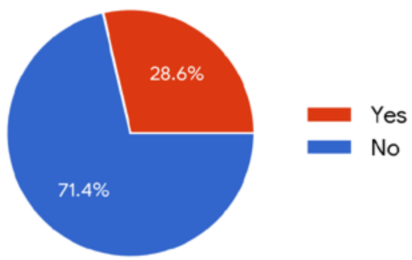


0.3 How often do you use AI for your core work tasks?
- ○ Never
- ○ Rarely
- ○ Sometimes
- ○ Often
- ○ Always (daily)



0.4 Have you received formal training on using AI in your work?
- ○ Yes
- ○ No

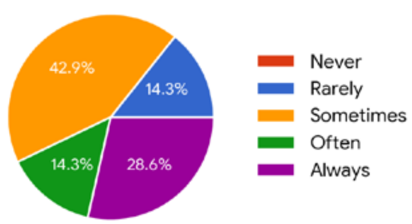
If yes, what kind of training? ____________



Individual Cognitive Mechanisms

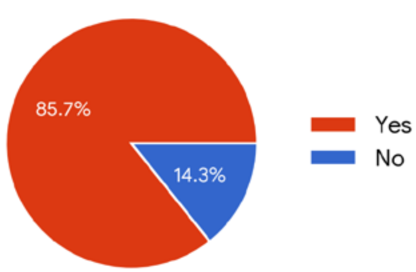
1.1 How often do you critically verify AI-generated outputs (e.g., code, calculations, drafts, recommendations) before using them?
- Never
- Rarely
- Sometimes
- Often
- Always

Notes: ____________________



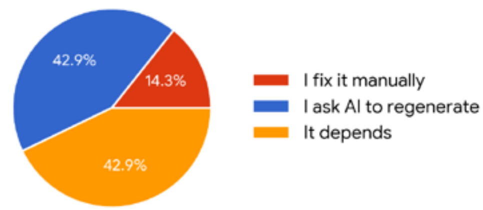
1.2 Do you ever delegate the exploratory phase of a task (e.g., identifying trends, brainstorming, scoping a problem) to AI, skipping the initial phase of forming your own approach?
- Yes
- No

Notes: ____________________



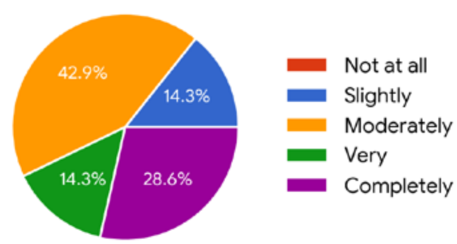
1.3 When the AI generates an output that contains an error, do you try to diagnose and fix it manually, or do you simply ask the AI to regenerate it?
- I fix it manually
- I ask AI to regenerate
- It depends

Notes: ____________________



1.4 How confident are you in your ability to perform basic, routine tasks (e.g., a manual check, a simple edit, a straightforward calculation) independently without AI assistance?
- Not at all
- Slightly
- Moderately
- Very
- Completely

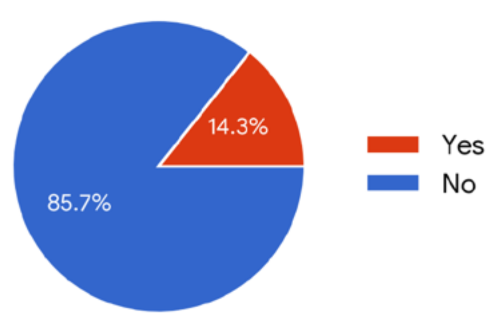
Notes: ____________________



Professional Development & Expertise Risks

2.1 Does your organization provide formal training on how to use AI to enhance core skills (e.g., prompt engineering, using AI as a learning/teaching tool)?
- Yes
- No

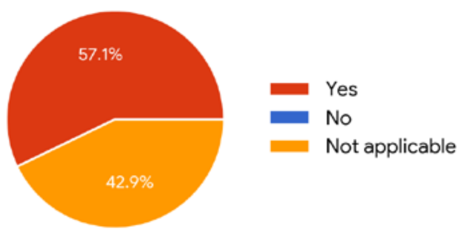
Notes: ____________________



2.2 Have you observed that junior professionals in your team are losing foundational skills (e.g., manual checks, reading raw logs, debugging simple scripts, understanding basic workflows) due to reliance on AI?

- Yes
- No
- Not applicable

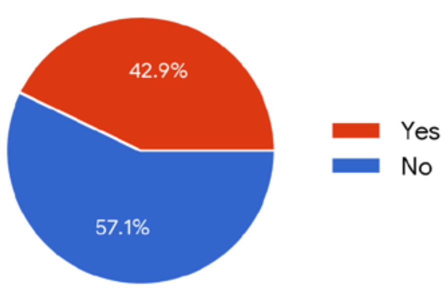
Notes: ______________________________



2.3 Do you have regular "No-AI" exercises (e.g., performing a task manually, writing a draft from scratch, validating a result) to verify genuine skill retention?

- Yes
- No

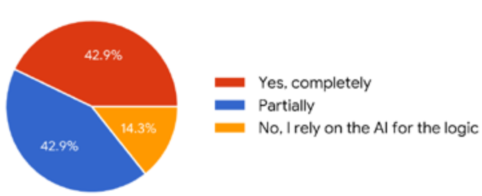
Notes: ______________________________



2.4 Can you explain the logic or reasoning behind a complex task that you have automated with AI, without needing the AI to re-explain it?

- Yes, completely
- Partially
- No, I rely on the AI for the logic

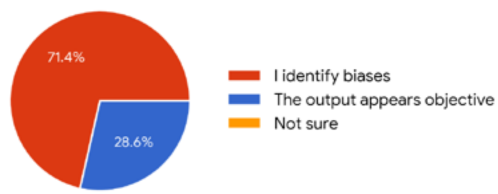
Notes: ______________________________



2.5 Are you able to identify potential biases or missing perspectives in an AI-generated output (e.g., a data summary, a list of recommendations, a draft document), or does the output always appear "objective" to you?

- I identify biases
- The output appears objective
- Not sure

Notes: ______________________________

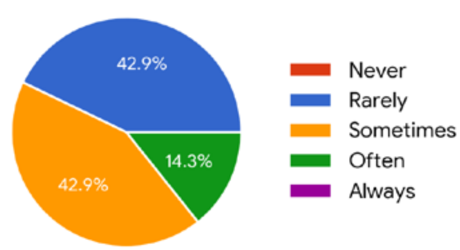


Structural & Organizational Factors

3.1 How often does your current workload allow enough time to critically review AI outputs before delivering final products or reports?

- Never
- Rarely
- Sometimes
- Often
- Always

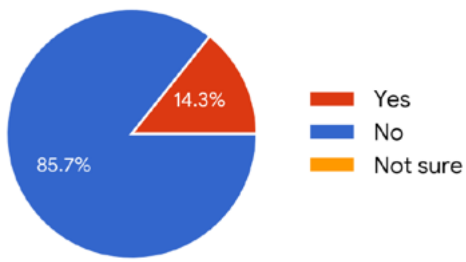
Notes: ______________________________



3.2 In your team, is there a documented "Plan B" for moments when the AI tool crashes, generates a wildly biased output, or hallucinates information?

- Yes
- No
- Not sure

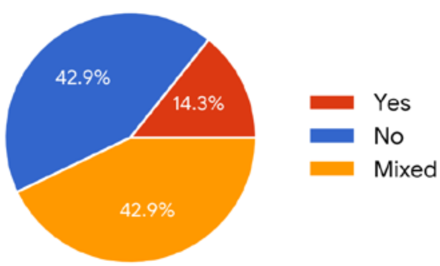
Notes: ______________________________



3.3 Does your organization incentivize accuracy and verification (e.g., rewards for catching AI errors) more than sheer output volume?

- Yes
- No
- Mixed

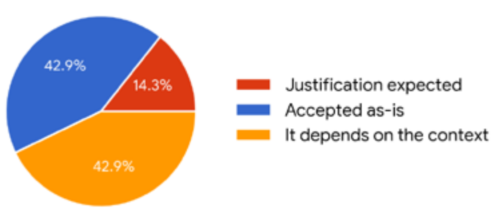
Notes: __________________________________



3.4 Are you expected to provide a justification and interpretation of AI-generated insights before they are included in a final output, or is the AI's output accepted as-is?

- Justification expected
- Accepted as-is
- It depends on the context

Notes: ______________________________________



Checklist V2.0

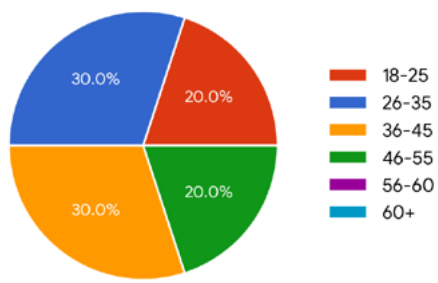
Deployment Metrics: *This intermediate version was compiled and field-tested by* ***10 End-User Testers****. It was subsequently reviewed by a new panel of* ***5 Expert Co-designers****, whose feedback on usability and scoring calibration shaped the final Version 3.0.*

0 — Participant Profile

0.1 What is your primary professional role?

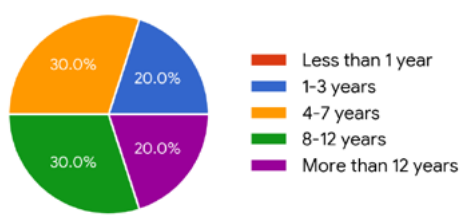
0.2 What age group do you belong to?

- 18–25
- 26–35
- 36–45
- 46–55
- 56–60
- 60+



0.3 How many years have you been working in your field?

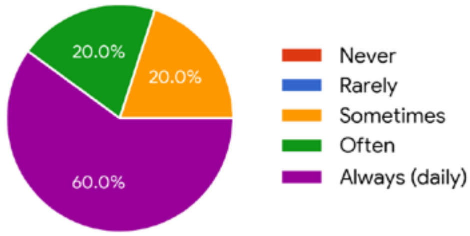
- Less than 1 year
- 1–3 years
- 4–7 years
- 8–12 years
- More than 12 years



0.4 How frequently do you rely on AI tools to complete your core work responsibilities?

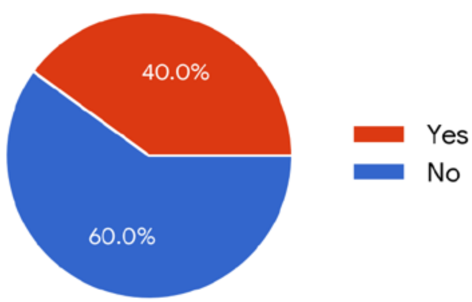
- Never
- Rarely
- Sometimes
- Often
- Always (daily)

If you rely on them, which tools do you use?______________



0.5 Have you received formal training on using AI in your work?
- Yes
- No

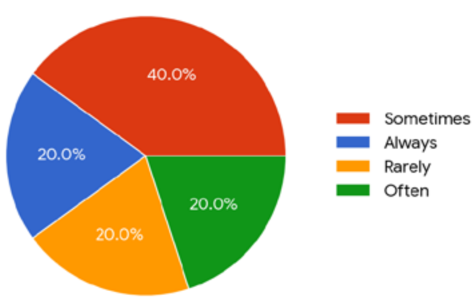
AI in your work? If yes, what kind of training?__________



1 — Individual Cognitive Mechanisms

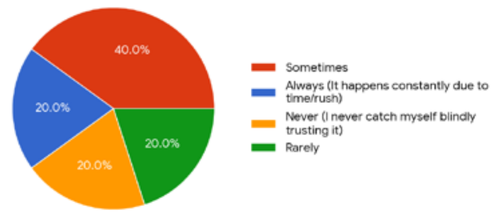
1.1 How often do you critically verify AI-generated outputs (e.g., code, calculations, drafts, recommendations) before using them?
- Never
- Rarely
- Sometimes
- Often
- Always

Notes:__________



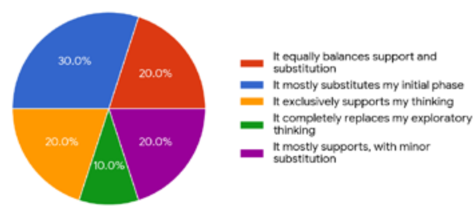
1.2 How often do you catch yourself passively trusting an AI output because it looks superficially correct, only to later realize it contained a hidden hallucination or error (e.g., invented data, wrong code)?
- Never (I never catch myself blindly trusting it)
- Rarely
- Sometimes
- Always (It happens constantly due to time/rush)

Notes:__________



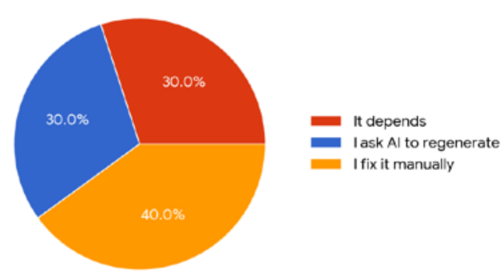
1.3 To what extent does the AI support the exploratory phase of a task (e.g., identifying trends, brainstorming, defining a problem) rather than substituting your own initial approach?
- It exclusively supports my thinking
- It mostly supports, with minor substitution
- It equally balances support and substitution
- It mostly substitutes my initial phase
- It completely replaces my exploratory thinking

Notes:__________



1.4 When the AI generates an output that contains an error, do you try to diagnose and fix it manually, or do you simply ask the AI to regenerate it?
- I fix it manually
- I ask AI to regenerate
- It depends

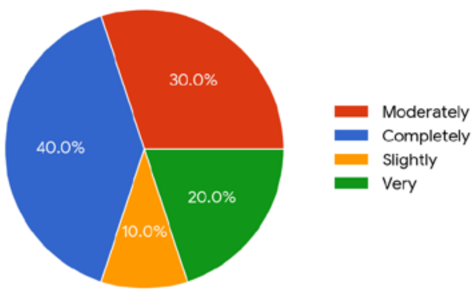
Notes:__________



1.5 How confident are you in your ability to perform basic, routine tasks (e.g., a manual check, a simple edit, a straightforward calculation) independently without AI assistance?
- Not at all

- Slightly
- Moderately
- Very
- Completely

Notes:________________________________

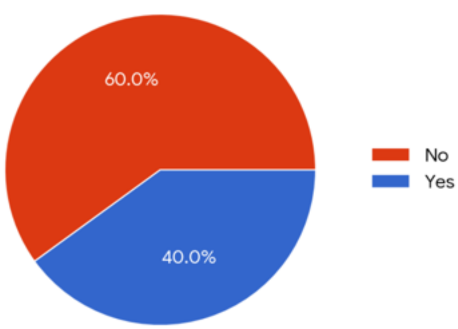


2 — Professional Development & Expertise Risks

2.1 Does your organization provide formal training on how to use AI to enhance core skills (e.g., prompt engineering, using AI as a learning/teaching tool)?

- Yes
- No

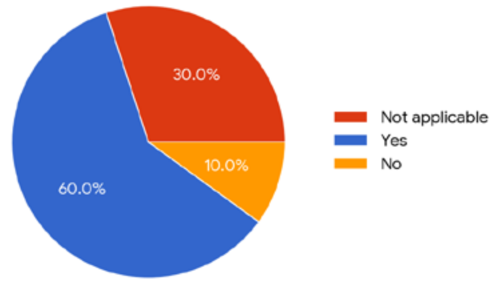
Notes:________________________________



2.2 Have you observed that junior professionals in your team are losing foundational skills (e.g., manual checks, reading raw logs, debugging simple scripts, understanding basic workflows) due to reliance on AI?

- Yes
- No
- Not applicable

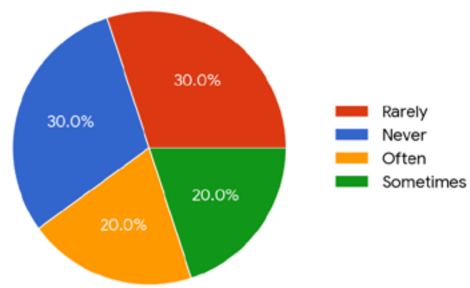
Notes:________________________________



2.3 Do you have regular "No-AI" exercises (e.g., performing a task manually, writing a draft from scratch, validating a result) to verify genuine skill retention?

- Never
- Rarely
- Sometimes
- Often
- Always

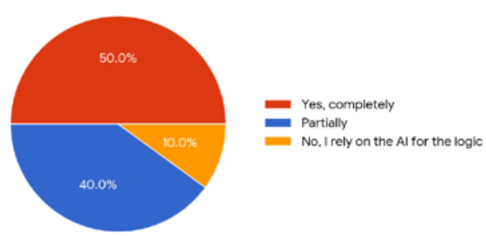
Notes:________________________________



2.4 Can you explain the logic or reasoning behind a complex task that you have automated with AI, without needing the AI to re-explain it?

- Yes, completely
- Partially
- No, I rely on the AI for the logic

Notes:________________________________

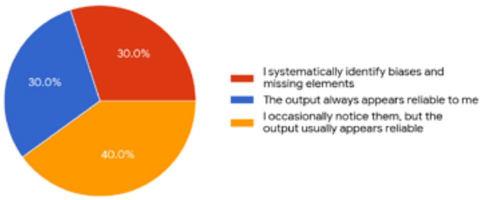


2.5 Are you able to identify potential biases or missing perspectives in an AI-generated output (e.g., a data summary, a list of recommendations), or does the output always appear "reliable" to you?

- I systematically identify biases and missing elements

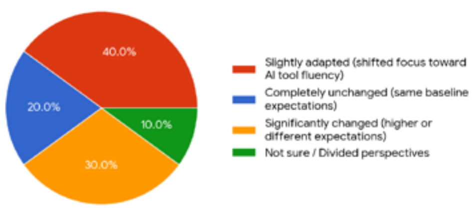
- I occasionally notice them, but the output usually appears reliable
- The output always appears reliable to me
- Not sure / I don't have enough domain expertise to tell

Notes:____



2.6 How has AI adoption changed the onboarding expectations or evaluation standards for new junior professionals within your team?
- Significantly changed (higher or different expectations)
- Slightly adapted (shifted focus toward AI tool fluency)
- Completely unchanged (same baseline expectations)
- Not sure / Divided perspectives

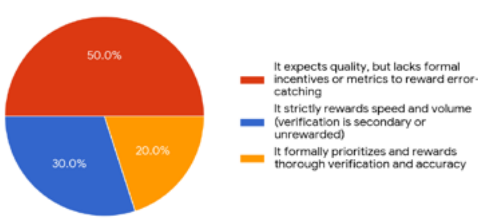
Notes:____



3 — Structural & Organizational Factors

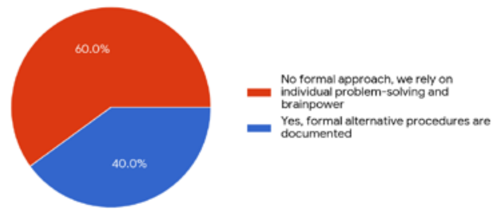
3.1 Does your organization structurally incentivize accuracy and verification of outputs, or does it primarily prioritize sheer work volume and speed?
- It strictly rewards speed and volume (verification is secondary or unrewarded)
- It expects quality, but lacks formal incentives or metrics to reward error-catching
- It formally prioritizes and rewards thorough verification and accuracy

Notes:____



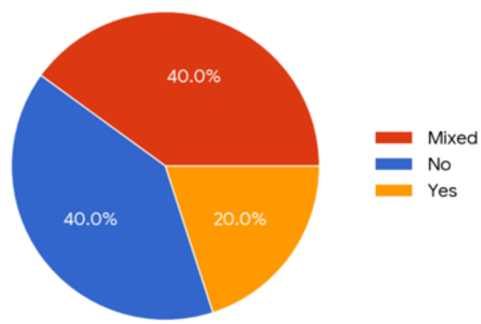
3.2 Are alternative operating procedures provided or documented within your team in the event of a malfunction or unreliability of AI tools?
- Yes, formal alternative procedures are documented
- No formal approach, we rely on individual problem-solving and brainpower
- Not sure

Notes:____



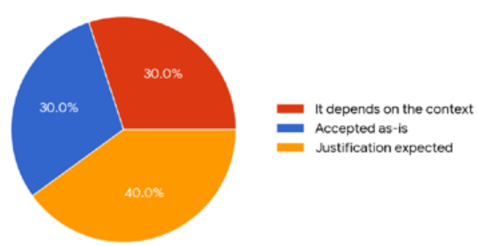
3.3 Does your organization incentivize accuracy and verification (e.g., rewards for catching AI errors) more than sheer output volume?
- Yes
- No
- Mixed

Notes:____



3.4 Are you expected to provide a justification and interpretation of AI-generated insights before they are included in a final output, or is the AI's output accepted as-is?
- Justification expected
- Accepted as-is
- It depends on the context

Notes:____



Final Question

Which core competency do you believe must be absolutely protected from erosion due to AI tools? *(select one or more)*

- Critical thinking and analytical skills (e.g., evaluating output quality, identifying logical errors, making inferences)
- Foundational technical skills (e.g., writing simple code, performing manual calculations, using basic tools)
- Domain or business knowledge (e.g., understanding the context, industry rules, client needs)
- Creativity and problem-solving (e.g., brainstorming, formulating hypotheses, exploring alternative solutions)
- Ethical judgment (e.g., evaluating social impact)
- Communication and collaboration skills (e.g., explaining results to non-experts, working in teams)
- Ability to learn rapidly (e.g., acquiring new skills without AI assistance)
- Other (please specify):____________________

Checklist V3.0 (Final Version)

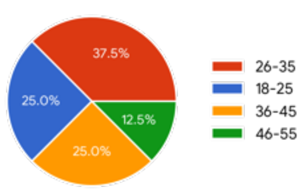
Deployment Metrics: *This final, optimized framework was compiled and field-tested by* ***8 End-User Testers****. Their empirical responses (P1–P8) were used to validate the final scoring matrix and are visualized in the Skill-Safe Framework Matrix in Section 6.*

0 — Participant Profile

0.1 What is your primary professional role? Write here: ______

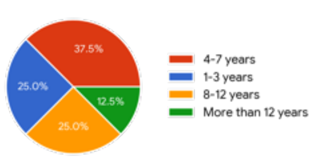
0.2 What age group do you belong to?

- 18–25
- 26–35
- 36–45
- 46–55
- 56–60
- 60+



0.3 How many years have you been working in your field?

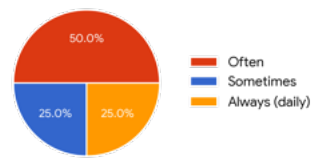
- Less than 1 year
- 1–3 years
- 4–7 years
- 8–12 years
- More than 12 years



0.4 How frequently do you rely on AI tools to complete your core work responsibilities?

- Never
- Rarely
- Sometimes
- Often
- Always (daily)

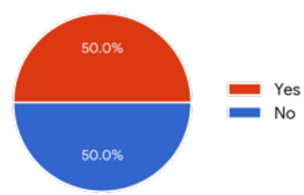
If you rely on them, which tools do you use? __________



0.5 Have you received formal training on using AI in your work?

- Yes
- No

If yes, what kind of training? ____________________

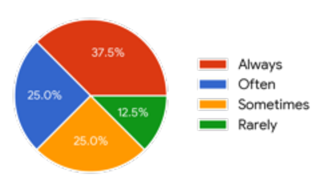


1 — Individual Cognitive Mechanisms

1.1 How often do you critically verify AI-generated outputs (e.g., code, calculations, drafts, recommendations) before using them?

- Never
- Rarely
- Sometimes
- Often
- Always

Notes (Optional): ____________________________

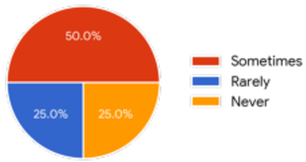


1.2 How often do you catch yourself passively trusting an AI output because it looks superficially correct, only to later realize it contained a hidden hallucination or error (e.g., invented data, wrong code)?

- Never (I never catch myself blindly trusting it)
- Rarely
- Sometimes

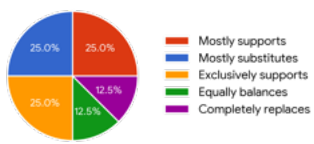
- Always (It happens constantly due to time/rush)

Notes (Optional): ________________



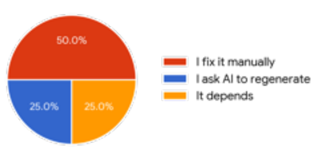
1.3 To what extent does the AI support the exploratory phase of a task (e.g., identifying trends, brainstorming, defining a problem) rather than substituting your own initial approach?
- It exclusively supports my thinking
- It mostly supports, with minor substitution
- It equally balances support and substitution
- It mostly substitutes my initial phase
- It completely replaces my exploratory thinking

Notes (Optional): ________________



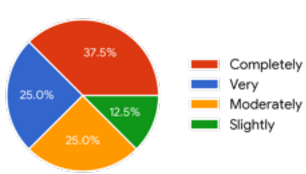
1.4 When the AI generates an output that contains an error, do you try to diagnose and fix it manually, or do you simply ask the AI to regenerate it?
- I fix it manually
- I ask AI to regenerate
- It depends

Notes (Optional): ________________



1.5 How confident are you in your ability to perform basic, routine tasks (e.g., a manual check, a simple edit, a straightforward calculation) independently without AI assistance?
- Not at all
- Slightly
- Moderately
- Very
- Completely

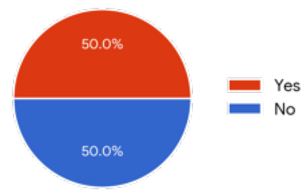
Notes (Optional): ________________



2 — Professional Development & Expertise Risks

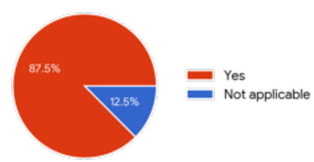
2.1 Does your organization provide formal training on how to use AI to enhance core skills (e.g., prompt engineering, using AI as a learning/teaching tool)?
- Yes
- No

Notes (Optional): ________________



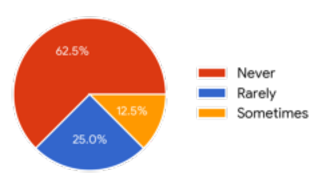
2.2 Have you observed that junior professionals in your team are losing foundational skills (e.g., manual checks, reading raw logs, debugging simple scripts, understanding basic workflows) due to reliance on AI?
- Yes
- No
- Not applicable

Notes (Optional): ________________



2.3 Do you have regular "No-AI" exercises (e.g., performing a task manually, writing a draft from scratch, validating a result) to verify genuine skill retention?
- Never
- Rarely
- Sometimes
- Often
- Always

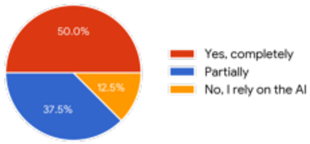
Notes (Optional): ________________



2.4 Can you explain the logic or reasoning behind a complex task that you have automated with AI, without needing the AI to re-explain it?
- Yes, completely

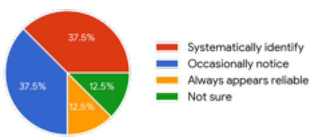
- Partially
- No, I rely on the AI for the logic

Notes (Optional): ____________________



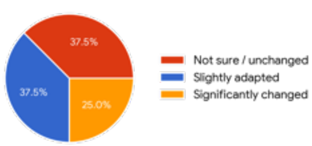
2.5 Are you able to identify potential biases or missing perspectives in an AI-generated output (e.g., a data summary, a list of recommendations), or does the output always appear "reliable" to you?
- I systematically identify biases and missing elements
- I occasionally notice them, but the output usually appears reliable
- The output always appears reliable to me
- Not sure / I don't have enough domain expertise to tell

Notes (Optional): ____________________



2.6 How has AI adoption changed the onboarding expectations or evaluation standards for new junior professionals within your team?
- Significantly changed (higher or different expectations)
- Slightly adapted (shifted focus toward AI tool fluency)
- Completely unchanged (same baseline expectations)
- Not sure / Divided perspectives

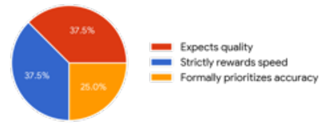
Notes (Optional): ____________________



3 — Structural & Organizational Factors

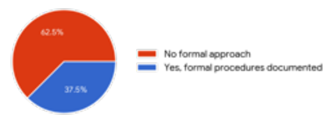
3.1 Does your organization structurally incentivize accuracy and verification of outputs, or does it primarily prioritize sheer work volume and speed?
- It strictly rewards speed and volume (verification is secondary or unrewarded)
- It expects quality, but lacks formal incentives or metrics to reward error-catching
- It formally prioritizes and rewards thorough verification and accuracy

Notes (Optional): ____________________



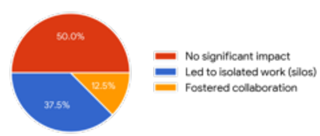
3.2 Are alternative operating procedures provided or documented within your team in the event of a malfunction or unreliability of AI tools?
- Yes, formal alternative procedures are documented
- No formal approach, we rely on individual problem-solving and brainpower
- Not sure

Notes (Optional): ____________________



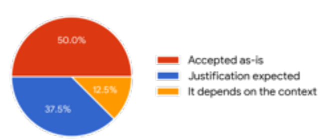
3.3 How has the integration of AI tools affected knowledge-sharing and peer-mentorship within your team?
- It has fostered collaboration and open discussion about AI limitations and solutions.
- It has had no significant impact on team communication and learning dynamics.
- It has led to isolated work (silos), reducing peer-to-peer learning and informal mentoring.

Notes (Optional): ____________________



3.4 Are you expected to provide a justification and interpretation of AI-generated insights before they are included in a final output, or is the AI's output accepted as-is?
- Justification expected
- Accepted as-is
- It depends on the context

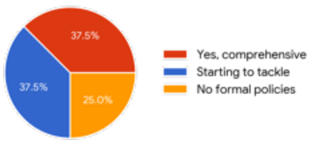
Notes (Optional): ____________________



3.5 Does your organization provide formal governance policies or review guidelines regarding the use of AI tools (e.g., addressing intellectual property issues or data privacy concerns)?

  - ◦ Yes, comprehensive governance policies are in place and enforced
  - ◦ No, but the company is starting to tackle these issues now
  - ◦ No formal policies exist
  - ◦ Not sure

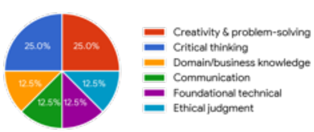
  Notes (Optional): ____________________



Final Question

Which core competency do you believe must be absolutely protected from erosion due to AI tools?
(Select one or more options)

- ◦ Critical thinking and analytical skills (e.g., evaluating output quality, identifying logical errors, making inferences)
- ◦ Foundational technical skills (e.g., writing simple code, performing manual calculations, using basic tools)
- ◦ Domain or business knowledge (e.g., understanding the context, industry rules, client needs)
- ◦ Creativity and problem-solving (e.g., brainstorming, formulating hypotheses, exploring alternative solutions)
- ◦ Ethical judgment (e.g., evaluating social impact)
- ◦ Communication and collaboration skills (e.g., explaining results to non-experts, working in teams)
- ◦ Ability to learn rapidly (e.g., acquiring new skills without AI assistance)
- ◦ Other (please specify): ____________________



Scoring and Feedback Guide

An Operational Metrics Framework to Assess Individual Deskilling Risks and Organizational Multipliers

Part I: Individual Autonomy Score (Behavioral Matrix)

To quantify your individual cognitive reliance on AI tools, assign points to your checklist selections from Section 1 and Section 2. This sub-score measures your personal autonomy, verification habits, and exposure to cognitive offloading risks.
Total Individual Autonomy Score: Sum the points from Items 1.1, 1.2, 1.3, 1.4, 1.5, 2.3, 2.4, and 2.5. (Minimum: 0 points, Maximum: 32 points).

Part II: Human-AI Interaction Archetypes

Match your total Individual Autonomy Score against the thresholds below to discover your behavioral user profile:

Profile A: The AI-Dependent Worker (Score: 0–10 points). High levels of cognitive offloading and automation complacency. The user routinely trusts algorithmic suggestions blindly and delegates problem-solving, facing severe risks of long-term expertise atrophy.

Profile B: The Vulnerable Synergist (Score: 11–21 points). Possesses task comprehension but frequently defaults to passive trust due to workload pressures or cognitive fatigue. Verification patterns are sporadic, introducing exposure to hidden AI hallucinations.

Profile C: The Autonomous Guardian (Score: 22–32 points). Excellent balance. AI is treated strictly as an assistant. Technical, analytical, and critical thinking skills are actively sustained through deliberate cognitive friction and rigorous manual validation.

Part III: Organizational Environment Score (Systemic Context)

To evaluate whether your workplace culture fosters or actively sabotages skill retention, convert your checklist responses from Section 3 into numerical modifiers using this operational matrix:
Total Organizational Score Calculation: Algebraically sum the modifiers (-1, 0, +1) chosen across items 3.1 to 3.5. (Minimum: -5 points, Maximum: +5 points).

Part IV: Environmental Archetypes

Map your total Organizational Score to identify your corporate ecosystem archetype and uncover its real-world impact on your skills:

The Erosion Trap (Score: -5 to -2 points). The workplace culture aggressively penalizes reflection. Systemic time pressures, misaligned metrics, and absent fallback infrastructures mean that even an autonomous practitioner will eventually be forced to succumb to cognitive offloading to survive project deadlines.

The Friction Zone (Score: -1 to +1 points). Institutional dissonance. The enterprise mandates manual accountability or justification, yet fails to provide the operational time window, training, or tools required. Workers experience severe cognitive fatigue attempting to manually safeguard quality against production friction.

The Skill-Safe Ecosystem (Score: +2 to +5 points). An idealized, highly mature framework. The organization actively values quality control, provides structural buffer times for manual code audits, enforces robust data governance, and rewards systemic anomaly detection, effectively preserving corporate intellectual capital.

Table 2: Individual Autonomy Score (Behavioral Matrix)

Item ID	Checklist Behavioral Question	Scoring Rubric & Point Values
Item 1.1	Critical Verification frequency of AI outputs before deployment.	Always (4) | Often (3) | Sometimes (2) | Rarely (1) | Never (0)
Item 1.2	Frequency of catching oneself passively trusting AI outputs.	Never (4) | Rarely (3) | Sometimes (1) | Always (0)
Item 1.3	Extent of AI support during the exploratory phase of a task.	Exclusively supports (4) | Mostly supports (3) | Equally balances (2) | Mostly substitutes (1) | Completely replaces (0)
Item 1.4	Approach to error correction (Diagnosis vs. Regeneration).	Manual fix (4) | It depends (2) | Ask AI to regenerate (0)
Item 1.5	Confidence in performing routine tasks without AI assistance.	Completely (4) | Very (3) | Moderately (2) | Slightly (1) | Not at all (0)
Item 2.3	Participation or availability of regular "No-AI" slots/exercises.	Always/Often (4) | Sometimes (2) | Rarely/Never (0)
Item 2.4	Ability to explain automated logic without AI assistance.	Yes, completely (4) | Partially (2) | No, I rely on the AI (0)
Item 2.5	Perception of AI summaries and bias detection capability.	I systematically identify biases (4) | Occasionally/Not sure (1) | Appears reliable/objective (0)

Table 3: Organizational Environment Score (Systemic Context)

Item ID	Checklist Organizational Question	Answer Mapping & Modifiers
Item 3.1	Does your organization incentivize verification of outputs, or does it primarily prioritize work volume?	Yes, completely (+1) Context-dependent / Partially (0) It strictly rewards speed (-1)
Item 3.2	Does your team have a documented "Plan B" (fallback workflow) if AI systems crash?	Yes, fully documented (+1) Informal/Not sure (0) No plan exists (-1)
Item 3.3	How has the integration of AI tools affected knowledge-sharing and peer-mentorship within your team?	It has fostered collaboration and open discussion. (+1) No significant impact / Neutral. (0) It has led to isolated work and silos. (-1)
Item 3.4	Are you expected to provide critical justification and interpretation of AI outputs?	Justification expected (+1) Context-dependent (0) Accepted as-is (-1)
Item 3.5	Does your organization provide formal AI governance policies (privacy, IP guidelines)?	Yes, enforced (+1) In progress / Not sure (0) No formal policies (-1)

Feedback from External Stakeholders (Phase 2 - Evolution from V1.0 to V2.0)

External Stakeholder ES1: CRM Data Analyst
Profile: CRM Data Analyst; uses AI every single day (Always); no official training courses; between 4 and 7 years of work experience.
Direct Quotes:

- On Section 3 (Corporate pressure and incentives): "It is very difficult to formally evaluate quality here. Errors are almost never measured or formally rewarded by the company. Theoretical accuracy is requested, but in reality, only delivery speed is incentivized."
- Global reflection gathered at the end of the session: "Which core competency do you think we absolutely need to protect if we keep delegating exploratory analysis?"

Design Implications for the V1 → V2 Transition:

1. Their strong feedback regarding the discrepancy between the theoretical requirement for accuracy and the actual push for speed convinced us to rework Item 3.1 in V2. We expanded it to understand whether or not the organization gives the worker enough actual time to perform a critical review before submitting a report.
2. Their final question about protecting skills felt like an excellent methodological point.

Because of this, we added the *Final Question* at the very bottom of V2 (a multiple-choice question to assess which core competencies, such as Critical thinking, are at the highest risk of eroding).

External Stakeholder ES2: AI Solution Architect
Profile: AI Solution Architect; daily use of AI (Always); practical on-the-job training; 1 to 3 years of experience in the role.
Direct Quotes:

- On framework terminology (Final discussion of V1): “Some terms used in the questionnaire, such as ’exploratory delegation’, are definitely too academic and risk confusing the respondents.”
- On internal corporate barriers: “The biggest issue is the lack of qualified resources who have the right mindset to face and manage such a drastic operational change.”

Design Implications for the V1 → V2 Transition:

1. Their comment forced us to carry out a major language-simplification process. Between V1 and V2, we cleaned up the text and added explanatory notes with ultra-practical examples (e.g., specifying that by "exploratory phase" we mean daily activities like brainstorming, scoping a problem, writing simple code).
2. To better capture the mental and corporate culture issues they mentioned, we introduced Meta-feedback boxes (Additions/Removals/Modifications) at the end of each section in V2, tied to traffic-light difficulty indicators (Green/Red).

External Stakeholder ES3: Bioengineering Researcher
Profile: Bioengineering Researcher (AI in Medicine); frequent use of AI (Often); self-taught; 4–7 years of experience. Represents the scientific core of the sample, highly focused on data accuracy and model bugs.
Direct Quotes:

- On model error mechanisms (Section 1): “...you have to be careful because generative model responses tend to mirror and adapt heavily to the user’s prompt. This creates plausible outputs that cater to human expectations, hiding underlying biases or hallucinations that are hard to spot if you don’t verify the source.”

Design Implications for the V1 → V2 Transition:

1. Their explanation of how models tend to "cater" to the user made us realize that V1 was too generic regarding cognitive biases. Therefore, in V2, we introduced Item 1.2 on latent Automation Bias ("How often do you catch yourself passively trusting an AI output because it looks superficially correct..."), explicitly linking this error to rush and time pressure ("due to time/rush").

External Stakeholder ES4: Senior Data Scientist
Profile: Senior Data Scientist; situational use of AI (Sometimes); self-taught; 4 to 7 years of professional experience. Represents the expert profile who already has a solid workflow and acts as a "shield" against deskilling.
Direct Quotes:

- On review habits (Item 1.1 of V1): “Personally, I only skip checking the results thoroughly when I am in a state of extreme rush or under intense pressure from management.”
- On skill evolution within the team: “AI usage doesn’t wipe out skills overnight, but it radically changes the onboarding process for juniors. Corporate performance expectations shift, and we risk losing the understanding of foundational theory.”

Design Implications for the V1 → V2 Transition:

1. Their reflections on the long-term impact within teams pushed us to shift our focus in Section 2 of V2, moving beyond the individual worker. We inserted Item 2.6 on professional skill adaptation to track how onboarding requirements and performance evaluations change when AI comes into play.
2. We also modified the options for Item 2.5 (Bias Detection) to clearly differentiate between users who passively assume the machine is reliable and those who actively manage to perform debiasing

External Stakeholder ES5: Junior Copywriter:
Profile: Junior Copywriter; daily use of AI (Always); no formal training; 1–3 years of experience in the sector. Represents the youngest bracket—digital natives who are nonetheless more exposed to the risk of losing procedural skills.
Direct Quotes:

- On response rigidity in the first version: “The strict choice between Yes and No in some questions of the first version cuts out all nuances. I often use AI not to replace my thinking, but as an iterative loop: it generates an error, I change the prompt, and it regenerates. It’s not a binary choice.”
- On open-ended text fields: “Having to write notes for every single question is painful; it makes you want to quit the survey halfway through.”

Design Implications for the V1 → V2 Transition:

1. This feedback directly led to a comprehensive reorganization of the visual layout of the multiple-choice options in V3. We strictly standardized the direction of all frequency scales: the lowest-risk scenario (Always for critical verification, or Never for blind trust) was consistently aligned on the exact same side across all items to prevent data-entry mistakes caused by visual fatigue (survey fatigue).

Feedback from External Stakeholders (Phase 3 - Transition from V2.0 to V3.0)

External Stakeholder ES6: Graphic Designer
Profile: Graphic Designer; uses AI very frequently (Often); 4–7 years of professional experience; no formal AI training.
Direct Quotes:

- On formatting and response alignment: “The response buttons for the multiple-choice questions keep changing their logical orientation from one section to another. Sometimes the positive or safe scenario is on the right and the risky one is on the left, and other times it is completely reversed. This lack of consistency makes it very easy to check the wrong box due to simple distraction during a quick reading.”

Design Implications for the V2 → V3 Transition:

1. This feedback directly led to a comprehensive reorganization of the visual layout of the multiple-choice options in V3. We

strictly standardized the direction of all frequency scales: the lowest-risk scenario (Always for critical verification, or Never for blind trust) was consistently aligned on the exact same side across all items to prevent data-entry mistakes caused by visual fatigue (survey fatigue).

External Stakeholder ES7: SEO Content Writer / Analyst
Profile: SEO Content Writer; daily use of AI tools (Always); 1–3 years of field experience; attended a corporate workshop on prompt engineering.
Direct Quotes:

- On redundancies within Section 3 (Organizational Factors): "Guys, Item 3.1 and Item 3.3 in V2 are basically asking the exact same thing using slightly different words. Both questions check whether the company rewards speed and volume over quality and error-checking. Filling out two identical questions just three lines apart feels like a pointless waste of time."

Design Implications for the V2 → V3 Transition:

1. To resolve this overlap identified in V2, we consolidated and removed the redundancy during the transition to V3. Item 3.3 was fully merged into Item 3.1, rephrasing the question so that the new unified item captures the organizational culture regarding accuracy in a much denser, more streamlined manner, thereby reducing the overall length of the section.

External Stakeholder ES8: Software Developer
Profile: Software Developer; daily use of AI tools (Always); 1–3 years of experience; self-taught.
Direct Quotes:

- On format inconsistencies (Item 1.3 and Section 2): "In the main report, we keep stating that we eliminated binary structures to avoid a superficial 'checkbox' mindset or 'compliance theatre'. However, if you look closely at V2, some old-school questions are still there: Item 1.3 starts with binary Yes/No radio buttons and then suddenly shows a 5-point scale right below it. The same goes for Items 2.1 and 2.2, which remained rigid Yes/No blocks. We need to be fully consistent with our own design philosophy!"

Design Implications for the V2 → V3 Transition:

1. This input was crucial for cleaning up the logical structure of the toolkit. In V3, we eliminated the format inconsistency in Item 1.3 by removing the opening Yes/No toggle and keeping only the nuanced behavioral frequency scale.
2. We completely removed the remaining dichotomous formats inherited from V1, converting all questions in Section 2 into graduated frequency scales to properly capture the real nuances of continuous training and skill maintenance habits.

External Stakeholder ES9: Data Analyst
Profile: Data Analyst; frequent use of AI tools (Often); 4–7 years of experience; no formal training.
Direct Quotes:

- On the absence of risk indicators and scoring: "The checklist gathers excellent data and definitely triggers reflection, but it is completely missing the most important feature for a manager or a team: a final output. Once I finish filling it out, what am I supposed to do with it? There is no score or benchmark that clearly tells me: 'Warning, the risk of skill erosion here is High, stop and provide training', or 'This process is Safe to delegate'."

Design Implications for the V2 → V3 Transition:

1. This was the defining methodological upgrade for V3. We formally introduced an interpretation guide and a comprehensive risk scoring system. In the final version of V3, the tool does not just collect answers; it provides a calculation framework (e.g., "If you score more than 3 answers within the critical tier in Section 2, a High Risk flag is triggered"), transforming the checklist into a genuinely actionable risk mitigation tool for organizations.

External Stakeholder ES10: HR Recruiter
Profile: HR Recruiter; frequent use of AI tools (Often); 4–7 years of field experience; attended a formal webinar on AI bias in hiring.
Direct Quotes:

- On the lack of temporal administration guidelines: "Deskilling and skill erosion are slow, gradual processes—they don't happen over a single afternoon. Yet, nowhere in this checklist does it explain how often an employee should fill it out, or at what specific point in the corporate workflow it should actually be deployed."

Design Implications for the V2 → V3 Transition:

1. This contribution allowed us to clarify the operational context of the evaluation framework. At the beginning of V3, we added a dedicated operational instructions section, explicitly clarifying that the checklist is not a one-off test. Instead, we specify that it must be administered periodically, recommending a regular cadence (such as every 6 months or during project retrospectives) to map and track risk trends over the long term.