



A Socio-Technical Systems Analysis Framework for Human-AI Collaborative Systems: A Case Study on AI-Powered Job Applicant Screening

Fawzeyia Elhashmi Mohamad Salah^{1*}, Nesrin Muftah Elrgibe²

^{1,2} Computer Department, Faculty of Education-Tripoli, Tripoli University, Libya.

Corresponding Author's Email: F.Salah@uot.edu.ly

Received: 08/9/2026 Revised: 10/9/2026 Accepted: 07/10/2026 Publication: 11/10/2026

Abstract

The increasing use of various types of probabilistic AI technologies in decision-making roles limits the usefulness of traditional business analysis and design (B&A) methods for documenting non-AI decision-makers and other aspects of probabilistic AI systems, including human backup decision-making roles. Traditional UML and other business analysis tools support only binary decision outcomes and offer few tools for documenting the shared decision-making responsibilities of AI and human decision makers. This paper proposes an AI-Socio B&A method that integrates traditional business analysis methods with three new tools: (1) a Confidence-Based Delegation Matrix offering three levels of AI decision-making authority (automatic decision making when confidence exceeds 80%, assisted decision making when confidence is between 50% and 80%, and human backup decision making when confidence is below 50%); (2) an amended use case format that offers additional documentation of confidence levels and bias-prevention techniques along with other information; (3) a process for obtaining and using feedback from human decision makers to improve the AI system. A case study of an AI-Powered Job Applicant Screening and Short-listing System, including documentation of an extended use case (UC-02), actors matrix, two design diagrams, and comparison with traditional B&A and UML documentation on five design and safety metrics, illustrates use of the new AI-Socio B&A method. Advantages of AI-Socio B&A include reduced automation bias and inclusion of anonymization as a requirement and other tools to share work between human and AI decision makers.

Keywords: Systems Analysis and Design (SAD), Human-AI Collaboration, Socio-Technical Systems, Job Applicant Screening, Extended Use Cases, Confidence Thresholds, Bias-Mitigation Gatekeepers.

المخلص

إن التزايد المستمر في استخدام تقنيات الذكاء الاصطناعي الاحتمالية المتنوعة في الأدوار المعنوية باتخاذ القرارات يحد من فائدة أساليب تحليل الأعمال وتصميمها التقليدية (B&A) في توثيق صانعي القرار غير المدعومين بالذكاء الاصطناعي، فضلاً عن جوانب أخرى لأنظمة الذكاء الاصطناعي الاحتمالية، بما في ذلك أدوار اتخاذ القرار الاحتياطية البشرية. وتقتصر أدوات النمذجة الموحدة (UML) وغيرها من أدوات تحليل الأعمال التقليدية على دعم النتائج الثنائية لاتخاذ القرار، كما أنها توفر أدوات محدودة لتوثيق مسؤوليات اتخاذ القرار المشتركة بين صانعي القرار من البشر وصانعي القرار المدعومين بالذكاء الاصطناعي. وتقتصر هذه الورقة أسلوباً تحليلياً يُعرف بـ "تحليل الأعمال الاجتماعي-التقني للذكاء الاصطناعي (AI-Socio B&A)"، يدمج أساليب تحليل الأعمال التقليدية مع ثلاث أدوات جديدة: (1) مصفوفة التفويض القائم على درجة الثقة (Confidence-Based Delegation Matrix) التي توفر ثلاثة مستويات من صلاحيات اتخاذ القرار للذكاء الاصطناعي (اتخاذ قرار آلي عندما تتجاوز درجة الثقة 80%، واتخاذ قرار بمساعدة بشرية عندما تتراوح درجة الثقة بين 50% و80%، واتخاذ قرار احتياطي بشري عندما تقل درجة الثقة عن 50%)؛ (2) صيغة معدلة لحالات الاستخدام تتضمن توثيقاً إضافياً لمستويات الثقة وتقنيات منع التحيز إلى جانب معلومات أخرى؛ (3) آلية للحصول على التغذية الراجعة من صانعي القرار البشر واستخدامها لتحسين نظام الذكاء الاصطناعي. دراسة حالة لنظام فرز وقوائم مختصرة للمتقدمين للوظائف المدعوم بالذكاء الاصطناعي متضمنة توثيق حالة استخدام موسعة، مصفوفة الفاعلين، مخططي تصميم، ومقارنة مع التوثيق التقليدي لكلا من B&A، UML وفق خمسة معايير للتصميم والسلامة؛

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: A Case Study on AI-Powered Job Applicant Screening

هذه الدراسة توضح طريقة AI-Socio B&A . ومن مزايا هذا الأسلوب الجديد : الحد من انحياز الأتمتة (Automation Bias)، وإدراج إخفاء الهوية (Anonymization) كمتطلب تصميمي، إضافة إلى أدوات أخرى لتوزيع العمل بين صانعي القرار من البشر والذكاء الاصطناعي. **الكلمات المفتاحية:** تحليل وتصميم النظم (SAD) ، التعاون بين الإنسان والذكاء الاصطناعي، النظم الاجتماعية-التقنية، فرز المتقدمين للوظائف، حالات الاستخدام الموسعة، عتبات الثقة، بوابات الحد من التحيز.

1. Introduction and Background

Software systems are undergoing a paradigm shift from deterministic, rule-driven artefacts to interactive, probabilistic systems in which artificial intelligence (AI) components act as semi-autonomous collaborators alongside human users. Berretta et al. (2023) observe that the increasing utilisation of AI requires a shift in perspective from technology-driven questions toward a human-centred research and design agenda that treats AI as more than a technological tool—as a teammate within socio-technical configurations. In a parallel development, Gonzalez et al. (2026) have argued that a key problem for human-AI interaction is not how to get humans and AIs to cooperate but rather how to make humans and AIs cooperate more effectively to achieve advantages of human-AI cooperation that neither individual humans nor individual AIs can achieve. The epistemological arguments of Gonzalez et al. have implications for the field of Systems Analysis and Design (SAD), because standard tools of SAD, including UML, use case, and other activity diagrams, were created under the assumption that systems will produce consistent outputs in response to inputs.

Traditional SAD methodologies, codified in canonical references such as Booch, Rumbaugh, and Jacobson (2005) and Sommerville (2015), model actors as either human users or deterministic external systems whose outputs are fully predictable. The use case template that accompanies these methodologies presupposes binary success or failure outcomes, encodes no notion of graded confidence, and offers no first-class constructs for representing fallback pathways triggered by model uncertainty. As a consequence, when contemporary generative and probabilistic AI components are inserted into enterprise workflows, classical SAD artefacts silently misrepresent the system: they hide the very uncertainty, bias, and delegation phenomena that determine whether the deployed system is safe, fair, and effective (Xu & Gao, 2026). The case of Amazon's experimental AI recruiting tool, scrapped in 2018 after it systematically downgraded female applicants because it was trained on male-dominated historical résumés (Dastin, 2018), illustrates how the omission of bias-mitigation constructs from upstream systems analysis propagates into catastrophic downstream harms.

1.1 Problem Statement

The main issue discussed in this article is the lack of suitable modeling languages, such as UML, to model non-deterministic actor behaviors and other aspects of human-AI collaboration designs. Fragiadakis, Diou, Kousiouris, and Nikolaidou (2025) report that research on human-AI collaboration faces significant barriers related to the characteristics of human-AI collaboration and limitations in existing requirements modeling approaches for representing human-AI collaboration. Schmager, Pappas, and Vassilakopoulou (2025) note

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

that there is no consensus on best practices for developing Human-Centered AI (HCAI) applications and no suitable modeling tools to support designers of HCAI applications who want to include human review in their designs. Other researchers have similarly noted a “double-edged sword” effect of using AI applications, including benefits such as increased productivity and harms such as discrimination and other risks resulting from automation bias and other factors when sociotechnical safeguards are not modeled during analysis and other phases of the systems development life cycle, according to Xu and Gao (2026).

1.2 Research Questions

Building on the foregoing gap, this study is guided by two research questions:

1. RQ1: How can traditional systems analysis models be extended to integrate AI agents as software “actors” with dynamic autonomy—such that graded confidence, fallback, and feedback phenomena become first-class modelling constructs rather than after-the-fact implementation details?
2. RQ2: What conceptual and technical fallback mechanisms are needed to mitigate bias or uncertainty risks in decision-support systems, and how can they be encoded into a use case template that is both rigorous enough for peer review and operational enough for real-world deployment?

1.3 Main Contributions

The paper advances three principal contributions. First, it formulates an extended socio-technical SAD framework—AI-Socio SAD—that augments classical use case modelling with confidence thresholds, model-uncertainty logic, and bias-mitigation gatekeepers. Second, it constructs a Confidence-Based Delegation Matrix formalising three levels of autonomy (auto-execution, human-in-the-loop, and fallback escalation) that correspond to empirically defensible confidence bands. Third, it applies the framework to a detailed case study of an AI-Powered Job Applicant Screening and Shortlisting System, producing a fully specified extended use case (UC-02), an actors matrix, two architectural visualisations, and a comparative evaluation against traditional UML/SAD across five quality and safety criteria. Together, these contributions operationalise the socio-technical design principle that Herrmann articulates, namely that AI adoption requires embedding intelligent components into communication, collaboration, and decision-making processes from a human-centred perspective.

2. Theoretical Framework: Socio-Technical SAD for AI Systems

The ideas in the proposed framework come from two traditions. One is traditional socio-technical systems theory, first introduced by Trist and Bamforth (1951) and elaborated further by Emery (1980). A core tenet of traditional socio-technical systems theory is that the performance of an organization's systems depends on the interaction of its social and technical elements, not on its social or technical elements individually. The second is the more recent literature on human-centred AI and human–AI teaming (Berretta et al., 2023;

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: A Case Study on AI-Powered Job Applicant Screening

Kargarnovin et al., 2026; Schmager et al., 2025), which extends STS thinking to configurations where one of the “teammates” is a probabilistic software agent. Xu and Gao (2026) propose an Intelligent Sociotechnical Systems (iSTS) framework that extends traditional STS theory to meet the demands of AI, emphasising human-centred joint optimisation across individual, organisational, ecosystem, and societal levels. The present paper translates that high-level framing into concrete SAD modelling constructs that can be inserted into existing requirements-engineering workflows.

2.1 Delegation Confidence Matrix

The core of AI-Socio SAD is the Confidence-Based Delegation Matrix, adapted from human-in-the-loop studies (AWS, 2024; Elementum, 2024) on how human reviewers can be used for various AI tasks depending on the confidence level associated with AI predictions. Proposed alternatives for delegating AI tasks to reviewers at different confidence levels include three options. Recommendations for human-AI task interaction at different confidence levels are conservative, particularly for high-risk predictions, based in part on studies by Kargarnovin et al. (2026) of the advantages of human-AI collaboration, such as improved coordination and communication, monitoring of AI tool use, and other benefits of mixed human and AI task performance at various points in a workflow.

The three levels are defined as follows:

Level	Confidence Band	Delegation Mode	Human Oversight Regime
Level 1	High confidence (> 80%)	Direct Full Automation (Auto-Execution)	None at decision time; logged for audit
Level 2	Moderate confidence (50–80%)	Supervisory Collaboration (Human-in-the-Loop)	HR Specialist reviews explainable fit report
Level 3	Low confidence (< 50%) or bias-risk flag	Human Review & Fallback (Escalation)	Hiring Manager manually reviews case

Table 1. Confidence-Based Delegation Matrix defining three autonomy levels for AI agents within the AI-Socio SAD framework.

The 80% upper threshold for default system decisions is used in other applications with confidence thresholds and is increasingly common in other applications as confidence thresholds are adopted or adapted for risk management purposes in addition to or instead of decision-making purposes (Research Square, 2024). The 50% lower threshold for human review is a conservative policy approach that assigns cases with less than 50% confidence in system decisions to human reviewers rather than the default system decision maker. Evidence from a meta-analysis by Vaccaro, Almaatouq, and Malone (2024) suggests that this conservative approach can help reduce errors made by human-AI teams on tasks such as reviewing job applications compared with errors made by humans or AI systems working alone.

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

Key point: Delegation matrices are not static lookup tools. Drawing on Holstein and Satzger's assertion that workers' mental models of their tasks shift when they use AI tools, delegation matrices can be used in a feedback cycle in which human reviewers' decisions are documented to improve the parameters and confidence intervals of the AI agent over time. These uses make delegation matrices dynamic learning tools within organizations, as Herrmann explains: “AI use requires and enables new forms of organizational learning, where specialists jointly interpret AI output, adapt workflows, and refine rules for system improvement.”

2.2 Augmented Use Case Model

The second theoretical contribution is an Extended Use Case Model that adds three new fields to the Cockburn-style use case documentation. The original Cockburn-style use case documentation, including actors, preconditions, postconditions, basic flow, alternative flows, and exception flows, is kept to maintain compatibility with other software analysis and design methodologies. Three new fields added to the Cockburn-style use case documentation are: (i) Confidence Thresholds, listing the confidence levels for each flow; (ii) Model Uncertainty Logic, explaining the steps the AI agent takes to handle model uncertainty; and (iii) Bias-Mitigation Gatekeepers, explaining techniques to filter out or alter sensitive data before it is fed into the AI agent. These contributions help answer RQ1 by introducing new terminology and documentation related to use cases.

The addition of a required field for Bias-Mitigation Gatekeepers is especially important. In traditional SAD, fairness and bias-mitigation considerations are handled separately from technical specifications in other documents and processes. The case of the Amazon recruitment tool (Dastin, 2018) illustrates the dangers of keeping these considerations and processes separate. Amazon's tool, which included information from entire job applicants' résumés with gendered words and discriminated against women on a large scale, was not designed with fairness in mind. Required anonymization and bias-mitigation fields for AI-Socio SAD are consistent with HCAI recommendations to consider social issues like fairness during the design process (Schmager et al., 2025; Xu & Gao, 2026).

3. Case Study Using Artificial Intelligence to Evaluate Job Applicants

To demonstrate the real-world use of AI-Socio SAD, the AI-Socio SAD framework was applied to the design of an AI resume-screening and candidate-selection system for a midsize company with knowledge workers. The intended purpose of the system is to assist human managers in selecting candidates to receive additional consideration. A case study on this topic was selected because it presents some of the same challenges often encountered using other SAD methods: uncertainty related to decision making and other selection decisions, including resume scoring; discrimination concerns, including collection of data about protected classes on resumes; and potential legal liabilities in making selections. As Kargarnovin et al. (2026) note, there is increasing but limited research on human-AI

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: A Case Study on AI-Powered Job Applicant Screening

interactions in high-stakes contexts. The case study described here offers one example to help other researchers and practitioners working in this area.

3.1 Actors Matrix

According to the AI-Socio SAD extended actor modeling method, the system has four actors defined by type (human or AI) and other factors, including level of autonomy and function. Table 2 provides an actors matrix.

Actor	Type	Autonomy Level	Primary Responsibility
Applicant	Human (external)	—	Submits CV / résumé via application portal
Screening AI Agent	AI Agent (probabilistic)	Dynamic (L1/L2/L3)	Parses CV, computes fit-score, generates explainable report
HR Specialist	Human (internal)	Supervisory (L2)	Reviews ambiguous cases, validates fit reports
Hiring Manager	Human (internal)	Escalation (L3)	Reviews fallback cases, makes final hiring decision

Table 2. Actors Matrix for the AI-Powered Job Applicant Screening System, distinguishing human actors from the probabilistic AI agent and annotating each with its autonomy level.

The core breakthrough in this actors matrix is the treatment of the Screening AI Agent not as a passive external system but as a full actor with dynamic autonomy. Traditional SAD would model the AI as a “system under design” with no agency of its own; AI-Socio SAD instead acknowledges, in line with Ju and Aral's (2026) field-experimental evidence that human–AI collaboration displays more delegation than human–human collaboration, that the AI agent actively shapes the workflow by deciding on a case-by-case basis which delegation level to invoke. This reframing has immediate architectural consequences: Workflow considerations would be to include a Confidence Evaluator step in the process to route cases and document the confidence and responsibility levels for each case handled.

3.2 Extended Use Case Specification: UC-02

This specification details UC-02, “Screen and Shortlist Job Applicant,” using the extended format. UC-02 is the primary use case. All three layers of delegation, the anonymization gatekeeper, explainable fit report, and feedback loop are used in this use case.

Use Case ID: UC-02

Use Case Name: Screenshot and Rank Job Applicant

Main Actor: Screening AI (Probabilistic and adaptive autonomous systems)

Secondary Actors: Applicant (external), HR Staff (L2 manager), Hiring Manager (L3 fallback)

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

Trigger: Applicant submits CV via the company's online recruitment portal.

Pre-conditions:

1. A job description with assigned skill weights and other job details has been imported.
2. The Screening AI Agent is trained. The confidence estimator model is trained.
3. An anonymizing gateway service exists with current policies for removing identifying information.
4. The Feedback & Decision Log Store is writeable.

Post-conditions: On success: The applicant gets automatic shortlisting on the company website, is added to the HR ticket queue, or is forwarded to the Hiring Manager. The decision feedback, including confidence score and other information, is stored in the Feedback Store.

On rejection: The rejection is logged in the system with a code and routed for further action, but not immediately rejected.

Extended Fields (AI-Socio SAD Additions):

- **Confidence Levels for Automatic vs. Assisted vs. Other Responses:** L1 automatic: 80% or higher confidence; L2 assisted: 50% to 80% confidence; L3 other: Less than 50% confidence or bias-risk flag = TRUE.
- **Uncertainty Modeling:** The Confidence Evaluator assigns a score to the candidate response using learned softmax weights for features related to skill similarity and other uncertainty signals about the input, such as parseability and familiarity with input type and word overlap between input and response. A separate risk score for potential bias will be assigned when other evidence of a protected characteristic is present in the CV and cleared by an Anonymization Gatekeeper.
- **Mitigation Technique:** Anonymization Gatekeeper (required architectural component): Anonymizes application data by removing identifying information such as the applicant's name, sex, image, and current residential and university addresses to prevent socioeconomic status inference before use by the AI. The anonymization gatekeeper is used in addition to and before the screening AI agent and on all system paths, including error and other backup paths.

Basic Workflow (with Anonymization Gatekeeper step)

1. The applicant submits a CV electronically through the recruitment website. The website assigns an ID number to the applicant.
2. A gatekeeper hides any identifying information about the CV applicant, including name, gender, picture, home, and previous university affiliation, from any AI application before the CV is given to an AI tool.
3. The Screening AI Agent evaluates a redacted CV and learns details about the candidate, such as skills and education, to assign an initial score based on the job ad.
4. The Confidence Evaluator generates a final confidence score $c \in [0, 1]$ using inputs including the fit score, estimate of epistemic uncertainty, and measure of potential bias.

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

5. The Confidence-Based Delegation Router evaluates c : If c is greater than 0.80, a referral is made to Level 1 (shortlist); if c is between 0.50 and 0.80, a referral is made to Level 2 (HR referral with a note describing fit); if c is less than 0.50 or bias-risk is TRUE, a referral is made to Level 3 (referral to Hiring Manager).
6. Automatically shortlist the applicant at Level 1 and document the action and rationale in the Feedback Store. Inform the applicant of the action.
7. At Level 2, the Explainable Fit Report Generator provides a detailed report about the candidate's skills and weaknesses. The report is shared with the HR Dashboard.
8. On Level 3, the Escalation Handler sends a candidate package (anonymized CV and other documents, test results, and information about test and other related biases) to the Hiring Manager.
9. All human reviewer decisions (approve, deny, request more information) are stored in the Feedback Store and shared with the Dynamic Weight Module to adjust the AI agent's weights over time.
10. The use case concludes when a final decision regarding an applicant being considered for an opportunity is made and provided to the applicant.

Alternative flow (Explainable Fit Report approach)

- A1.** Triggers an alarm if the Confidence Evaluator gives a rating between 0.50 and 0.80 (Level 2).
- A2.** The Explainable Fit Report Generator reports the top- k skills the candidate possesses and the top- k missing skills, along with measures of uncertainty, and provides a report viewable by the Hiring Manager on the HR Dashboard.
- A3.** HR Specialist views the report and assesses the strengths and weaknesses analysis provided by the AI. The HR Specialist may confirm or disagree with the AI's recommendation and seek additional information from the Applicant.
- A4.** The HR Specialist's decisions are entered into the Feedback Store along with the version of the job report they reviewed. This information serves as feedback for use in future recalibration rounds.
- A5.** Optional text entry field to enable HR Specialists to comment on AI recommendations for audit and other purposes when they disagree with AI recommendations.
- A6.** Resume processing at Basic Flow step 9.

Fallback and Error Handling (unknown document type, lexical bias)

- E1.** Fired in cases where (a) file corruption or other problems prevent initial screening tools from successfully interpreting CVs, (b) other screening tools detect indicators of protected class membership that were not eliminated by anonymization, or (c) $c < 0.50$.
- E2.** The Escalation Handler creates an escalation package that includes an anonymized CV, parsed CV, uncertainty data, and a structured escalation code (UNKNOWNFORMAT, BIASFLAG, LOW_CONFIDENCE).

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

E3. The packet is high priority in the Hiring Manager's inbox. Automated rejection is prohibited.

E4. The Hiring Manager evaluates the application materials and may ask the Applicant to resubmit materials in a different format. The Hiring Manager will make a decision at this stage.

E5. The decision, reason code, and (optional) rationale are written to the Feedback Store; recurring UNKNOWN_FORMAT or BIAS_FLAG patterns trigger an alert to the system administrator to update the parser or the Anonymization Gatekeeper rules.

E6. Control returns to Basic Flow step 9.

4. Architectural Visualizations and System Diagrams

This section presents two architectural visualisations that operationalise the AI-Socio SAD constructs developed above. Figure 1 is an Extended Activity Diagram that illustrates the decision-flow paths, the confidence-branching logic, and the human-fallback nodes. Figure 2 is a Human–AI Interaction Architecture Flow Diagram depicting the principal system components and the feedback loop that channels human decisions back into AI recalibration. Both figures are accompanied by their renderable Mermaid source code so that practitioners can reproduce or adapt them in standard modelling tools.

4.1 Figure 1: Extended Activity Diagram

Figure 1 traces a single applicant's CV through the full screening pipeline. The Anonymization Gatekeeper runs first, ensuring that no AI component ever ingests raw sensitive attributes. The Screening AI Agent and Confidence Evaluator then jointly produce a confidence value, which is routed by the central decision diamond into one of three branches corresponding to the three delegation levels of Table 1. Crucially, all three branches converge on a common Decision Logging and Feedback Store, from which a dashed feedback arc returns to the AI Agent—visualising the dynamic weight-adjustment loop. The diamond and the three coloured branch nodes are precisely the modelling primitives that traditional UML activity diagrams lack.

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems:
A Case Study on AI-Powered Job Applicant Screening

Figure 1. Extended Activity Diagram for AI-Powered Job Applicant Screening
(Decision Flow, Confidence Branching, and Human Fallback Nodes)

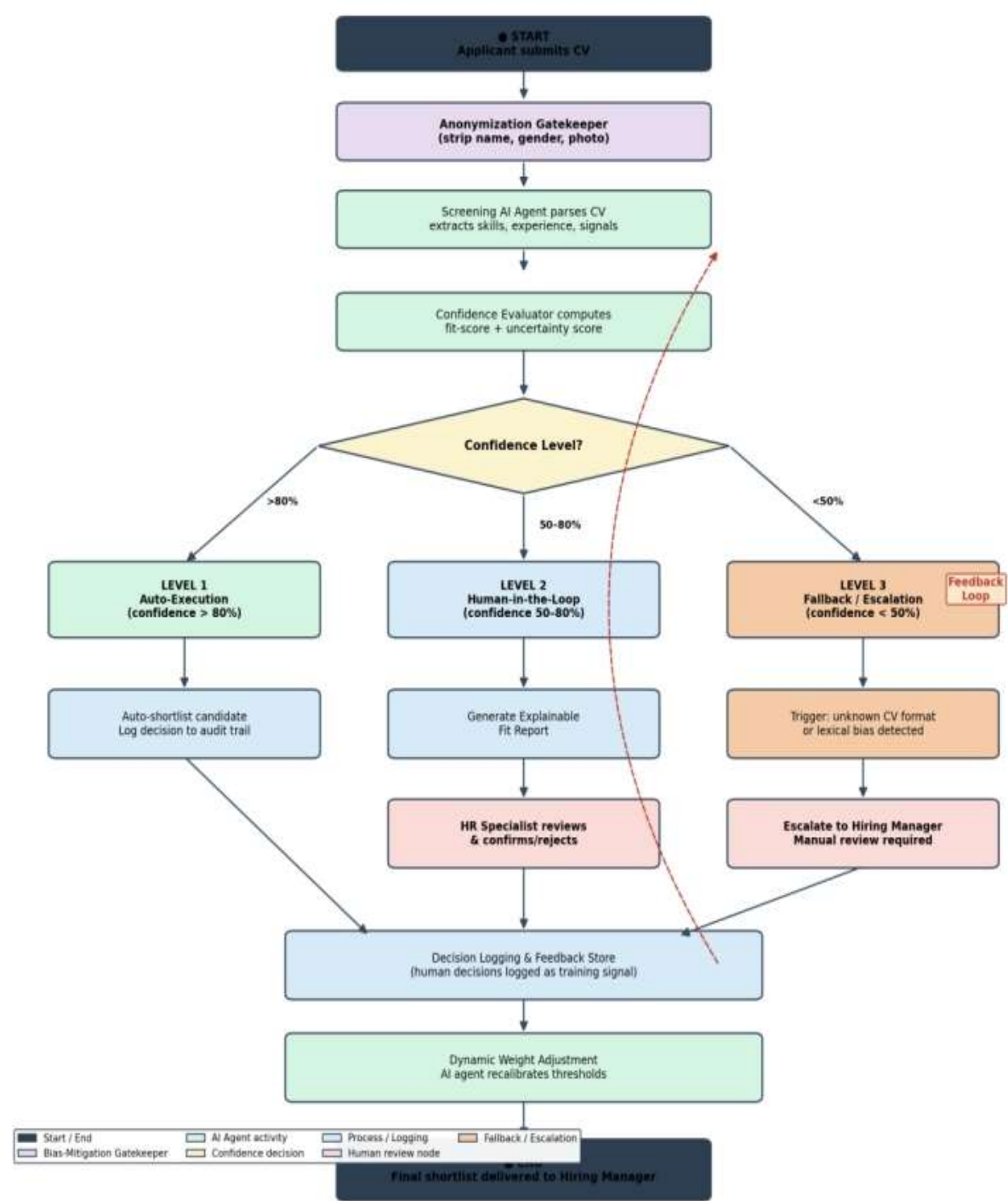


Figure 1. Extended Activity Diagram for AI-Powered Job Applicant Screening, showing confidence-branching decision flow and human-fallback nodes.

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

The renderable Mermaid source for Figure 1 is provided below for reproduction in standard tooling:

flowchart TD

```
A([START: Applicant submits CV]) --> G[Anonymization Gatekeeper<br/>strip name, gender, photo]
G --> P[Screening AI Agent parses CV<br/>extracts skills, experience, signals]
P --> C[Confidence Evaluator computes<br/>fit-score + uncertainty]
C --> D{Confidence Level?}
D -->|> 80%| L1[LEVEL 1: Auto-Execution]
D -->|50-80%| L2[LEVEL 2: Human-in-the-Loop]
D -->|< 50% or bias| L3[LEVEL 3: Fallback / Escalation]
L1 --> S1[Auto-shortlist candidate<br/>Log decision to audit trail]
L2 --> R[Generate Explainable Fit Report]
R --> H1[HR Specialist reviews<br/>and confirms / rejects]
L3 --> FB[Trigger: unknown format<br/>or lexical bias detected]
FB --> H2[Escalate to Hiring Manager<br/>Manual review required]
S1 --> LOG1[(Decision Logging &<br/>Feedback Store)]
H1 --> LOG1
H2 --> LOG1
LOG1 --> W[Dynamic Weight Adjustment<br/>AI agent recalibrates thresholds]
W --> END([END: Final shortlist delivered])
LOG1 -. Feedback Loop .-> P
```

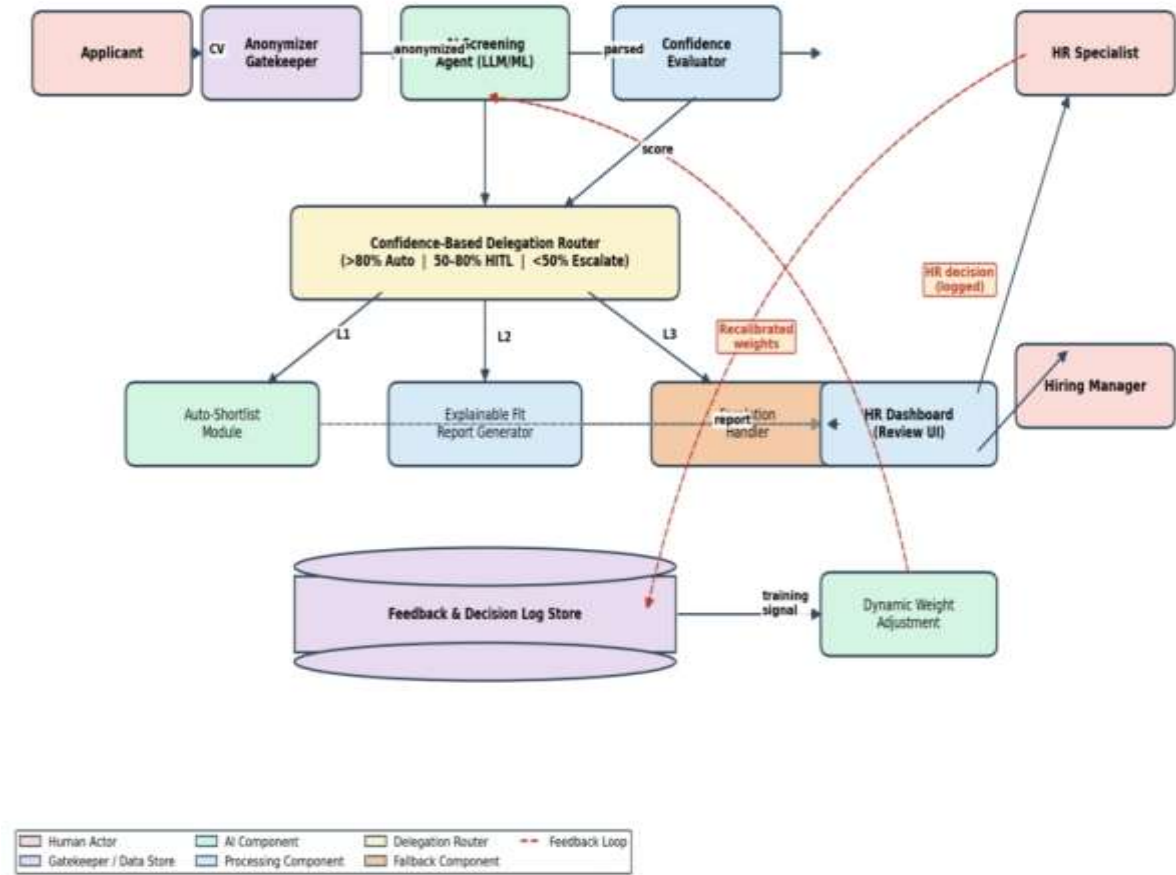
4.2 Figure 2: Human-AI Interaction System Diagram

Figure 2 shows a diagram of the screening system. Five elements of the system (Anonymizer, AI Screening Agent, Confidence Assessment Agent, HR Portal, and Feedback and Outcome Database) and five human users of the system (Applicant, HR Representative, Hiring Manager) and the Confidence-Based Delegation Algorithm are shown. Dashed lines in the diagram represent feedback to the system involving human decision making that is documented in the Feedback and Outcome Database and used by the Dynamic Weight Adjustment Algorithm to improve the AI Screening Agent, as discussed by Herrmann.

**A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems:
A Case Study on AI-Powered Job Applicant Screening**

Figure 2. Human–AI Interaction Architecture Flow Dia

Figure 2. Human-AI Interaction Architecture Flow Diagram
(System Components: Anonymizer, AI Agent, Confidence Evaluator, HR Dashboard, Feedback Loop)



gram showing system components (Anonymizer, AI Agent, Confidence Evaluator, HR Dashboard) and the feedback loop that recalibrates the AI agent from logged human decisions.

The renderable Mermaid source for Figure 2 is provided below:

```
graph LR
    APP([Applicant]) --> ANON[Anonymizer Gatekeeper]
    ANON --> AI[AI Screening Agent<br/>LLM / ML]
    AI --> CE[Confidence Evaluator]
    CE --> ROUTER{Confidence-Based<br/>Delegation Router}
    ROUTER -->|L1: > 80%| AUTO[Auto-Shortlist Module]
    ROUTER -->|L2: 50-80%| EFR[Explainable Fit<br/>Report Generator]
    ROUTER -->|L3: < 50% or bias| ESC[Escalation Handler]
    AUTO --> DASH[HR Dashboard<br/>Review UI]
    EFR --> DASH
    ESC --> DASH
    DASH --> HRS([HR Specialist])
    DASH --> HM([Hiring Manager])
    HRS -->|HR decision logged| STORE[(Feedback &<br/>Decision Log Store)]
    HM -->|HM decision logged| STORE
    STORE --> W[Dynamic Weight Adjustment]
    W -->|Recalibrated weights| AI
```

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: A Case Study on AI-Powered Job Applicant Screening

5. Results and Evaluation

To assess the benefits of the AI-Socio SAD approach over other UML/SAD approaches, we conducted a comparative assessment using five quality and safety measures commonly used in human-centered AI and HAIC assessments (Fragiadakis et al.; Schmager et al., 2025; Xu & Gao, 2026). The five measures were applied to all assessed methods. Table 3 compares the representational capabilities of current methods and new methods. Results of applying new methods to a job evaluation task are shown in Table 3. Other factors evaluated in Table 3 include various socio-technical factors: actor representation, risk analysis, bias mitigation, transparency, and learning or feedback mechanisms.

Evaluation Criterion	Traditional Analysis (Traditional UML/SAD)	Proposed Framework (AI-Socio SAD)	Impact / Improvement in Case Study (Job Screening)
Actor Modeling	Human actor or deterministic external system.	AI Agent with dynamic autonomy and variable delegation levels.	Enables AI agents to make automated screening decisions under safety constraints.
Uncertainty Management	Assumes binary success/failure outcomes.	Models automated fallback pathways when model confidence drops.	Prevents unqualified auto-rejections caused by CV ambiguity or formatting issues.
Bias Mitigation	Handled outside documentation (not integrated into diagrams).	Integrated as a mandatory architectural gatekeeper (Anonymization Gatekeeper).	Automatically strips sensitive personal attributes (e.g., name, gender) prior to screening.
Decision Explainability	Final outputs without documented reasoning for uncertainty.	Generates Explainable Fit Reports for human reviewers.	Provides clear summaries of candidate strengths and skill gaps to the HR Specialist.
Learning & Feedback Loop	Manual reconfiguration via system re-programming.	Logs human reviewer decisions to dynamically adjust AI agent weights.	Continuously improves screening accuracy based on HR decision patterns.

Table 3. Comparative evaluation of Traditional UML/SAD against the proposed AI-Socio SAD framework across five quality and safety criteria, with concrete impacts observed in the job-screening case study

A comparative analysis shows several advantages of using AI elements in SAD. Advantages of using AI elements in SAD include removing assumptions about human-AI interaction because control delegation relationships are no longer modeled. Explicitly modeling control delegation in AI-Socio SAD helps analysts analyze and enhance human-AI interaction. Other advantages include explicitly modeling uncertainty in AI decision making to prevent rejection of CVs with unusual formatting, which has been a problem with some deployed AI hiring tools (Dastin, 2018), and requiring inclusion of bias-mitigation strategies in the design process to address fairness concerns, as suggested by the iSTS project to integrate

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

sociotechnical considerations into the design process rather than address them after deployment (Xu & Gao, 2026).

Results such as those presented in this paper are consistent with other research studies. A paper by Vaccaro et al. (2024) finds that, on several decision-making tasks, average outcomes for human-AI decision-making teams are worse than for humans or AI making decisions independently. The outcomes for human-AI decision-making teams improve if the team uses a process similar to the one described in this paper in which low-confidence decisions are referred to a human decision maker rather than an AI decision maker.. Ju and Aral (2026) find that human–AI collaboration displays more delegation and more task-oriented communication than human–human collaboration; the Confidence-Based Delegation Matrix provides the modelling vocabulary to design for, rather than against, this delegation behaviour. Fragiadakis et al. argue that HAIC evaluation must integrate both quantitative and qualitative criteria; Table 3 operationalises this integration by juxtaposing representational capability (qualitative) with case-study impact (quantitative/empirical).

6. Discussion and Ethical Considerations

6.1 Mitigation of Automation Bias

Automation bias, in which human decision makers increasingly defer to algorithmic suggestions, is an important issue for decision-support interventions (Kargarnovin et al., 2026; Schmager et al., 2025). AI-Socio SAD includes strategies to reduce automation bias. The Confidence-Based Delegation Matrix ensures that AI-generated suggestions will be automatically provided only when the confidence level exceeds 80%; lower or moderate confidence level suggestions are provided to human decision makers so that human decision makers are less likely to view AI-generated suggestions as default recommendations. The Explainable Fit Report includes information about the AI's decision-making process, such as the best and worst skill-job matches and other diagnostics, as recommended by Holstein and Satzger to promote the formation of new mental models among decision makers. A logging requirement for human decision makers' override decisions provides information to managers to help them address instances of automation bias and prevent the bias from becoming a routine practice.

It is important to recognise, however, that automation bias cannot be eliminated by modelling alone. Gonzalez et al. (2026) discuss sociotechnical factors, such as trust and other motivational factors at the organizational level, that can help ensure human-AI complementarity. AI-Socio SAD is an important but not complete approach to safely implementing human-AI systems. It provides a starting point that other organizational approaches can build on to promote human-AI complementarity.

6.2 Limitations

Three limitations of the study are noted. The 80% and 50% autoexecute thresholds described in this study can be used in other contexts without change but may need to be tailored to suit

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

the needs of a particular employer. A 90% autoexecute threshold may be suitable for hiring senior executives to reduce costs from false positives, while lower autoexecute thresholds may be suitable for hiring large numbers of junior staff. Autoexecute thresholds should be set in collaboration with human resources and other internal stakeholders. Second, this work is more of an analysis and planning paper than an empirical study, and there is limited evidence on the effectiveness of the proposed framework in actual settings over time. According to a meta-analysis by Vaccaro et al. (2024), individuals paired with AI often have lower productivity than either individuals or AI alone, suggesting a need for further empirical research before widespread adoption of the framework. Third, the Anonymization Gatekeeper is a component of the proposed framework that does not fully prevent discrimination, as synonyms for protected groups can survive the anonymization process, among other issues related to training data and job descriptions and performance metrics used in the framework. The anonymization process is one tool to combat discrimination in the proposed framework.

6.3 Ethical Considerations

In addition to automation bias, three other ethical considerations should be noted. Ethical considerations about documenting human decisions in the feedback loop include privacy and surveillance considerations. Second, the Explainable Fit Report, while beneficial for transparency, can itself become a vector for discrimination if its “skill gap” framing systematically disadvantages candidates from non-traditional backgrounds; report templates should be audited for framing effects. Third, the dynamic recalibration of the AI agent from human decisions encodes a value judgement that “the human is right”—a judgement that may perpetuate historical bias if human reviewers themselves are biased. Schmager et al. (2025) and Xu and Gao (2026) both emphasise that human-centred AI is not equivalent to human-unchallenged AI; the feedback loop should therefore incorporate bias-detection monitors that flag systematic divergences between human and AI judgements on protected-attribute subgroups.

7. Conclusions and Future Work

This paper presents AI-Socio SAD, a socio-technical systems analysis approach that extends traditional UML-based SAD approaches in multiple ways with new methods and tools, including the Confidence-Based Delegation Matrix with three plausible degrees of autonomy, the Extended Use Case Model with additional tools and features such as confidence thresholds and model-uncertainty and bias gatekeeper analysis, and tools to help integrate human decisions into the tuning process for AI agents. An extended case study using the AI-Socio SAD approach to design an AI-Powered Job Applicant Screening and Shortlisting System produced deliverables including a full extended use case (UC-02), actors matrix, two diagrams, and a comparison of the AI-Socio SAD and traditional UML and SAD approaches based on five criteria. The assessment indicates that AI-Socio SAD addresses some of the shortcomings of traditional SAD approaches that limit their applicability to

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

probabilistic human-AI systems: Dynamic delegation and contingency planning, partial trust, and other capabilities can be incorporated into the AI-Socio SAD approach but are often included as add-on features in traditional SAD approaches.

The framework provides two key scientific advantages. Theoretically, the framework provides methods for incorporating sociotechnical approaches discussed in HCAI and iSTS literature (Herrmann; Schmager et al., 2025; Xu & Gao, 2026) into requirements specifications. Practically, the framework provides a template with broadly applicable tools, including a delegation matrix, anonymization gatekeeper, explainable output, and feedback loop, that other organizations can use. Although the case study example is a job application evaluation tool, the framework could be used for other tools, such as financial evaluation tools (e.g., credit evaluation tools with anonymization gatekeepers that block information about race and postal code) and clinical and administrative evaluation tools (e.g., diagnostic tools with confidence gates that send patients to physicians for further evaluation and benefit evaluation tools with escalation processes that prevent automatic denial of benefits to vulnerable applicants). Benefits of using the framework for these tools are similar to those discussed for the job application evaluation tool. It helps analysts communicate effectively with nontechnical users about issues of delegation and fairness, which are implicit in traditional SAD approaches.

Three areas for future research are identified. Methods should be longitudinally tested in field applications to verify the advantages of the framework, including reduced automation bias, and provide information about appropriate confidence levels for different types of projects. Also, tools can be developed to make the framework more accessible to others. Possible tools include an add-on for popular UML modeling software products such as Visual Paradigm and Enterprise Architect and open-source products that include preformatted Extended Use Case fields and other features to streamline creation of a delegation matrix according to specified confidence level inputs from users. Mathematical verification methods described in Research Square (2024) can be integrated with the framework to provide other advantages, including better control over residual errors introduced by automation. Integrating mathematical verification methods can also help bridge the gap between the practices discussed in this paper and formal methods and provide other benefits for systems analysis and design, such as a greater focus on users.

8. References

- AWS. (2024). What is human in the loop? HITL in machine learning. Amazon Web Services. <https://aws.amazon.com/what-is/human-in-the-loop>
- Berretta, S., Tausch, A., Ontrup, G., Gilles, B., Peifer, C., & Kluge, A. (2023). Defining human–AI teaming the human-centered way: A scoping review and network analysis. *Frontiers in Artificial Intelligence*, 6, 1250725. <https://doi.org/10.3389/frai.2023.1250725>
- Booch, G., Rumbaugh, J., & Jacobson, I. (2005). *The Unified Modeling Language user guide* (2nd ed.). Addison-Wesley.
- Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG>
- Elementum. (2024). Human-in-the-loop workflows: A complete guide. <https://www.elementum.ai/blog/human-in-the-loop-workflows>

A Socio-Technical Systems Analysis Framework for Human–AI Collaborative Systems: *A Case Study on AI-Powered Job Applicant Screening*

- Emery, F. E. (1980). Educational paradigms. In A. H. Burrell & K. E. Trist (Eds.), *The social engagement of social science* (Vol. 1, pp. 170–198). University of Pennsylvania Press.
- Fragiadakis, G., Diou, C., Kousiouris, G., & Nikolaidou, M. (2024). Evaluating human–AI collaboration: A review and methodological framework. Department of Informatics & Telematics, Harokopio University of Athens.
- Gonzalez, C., Donahue, K., Goldstein, D. G., Heidari, H., Jalali, M. S., Schelble, B., Singh, A., & Woolley, A. W. (2026). Toward a science of human–AI teaming for decision making: A complementarity framework. *PNAS Nexus*, 5, pgag030. <https://doi.org/10.1093/pnasnexus/pgag030>
- Herrmann, T. (2024). Organizational practices and socio-technical design of human-centered AI. In *Handbook of human-centered artificial intelligence*. Springer. https://doi.org/10.1007/978-981-97-8440-0_66-1
- Holstein, J., & Satzger, G. (2025). Development of mental models in human–AI collaboration: A conceptual framework [Completed research paper]. Karlsruhe Institute of Technology.
- Ju, H., & Aral, S. (2026). Collaborating with AI agents: A field experiment on teamwork, productivity, and performance. arXiv:2503.18281. <https://arxiv.org/abs/2503.18281>
- Kargarnovin, S., Hernandez, C. I., Reiners, D., Cruz-Neira, C., Bochenek, G., & Karwowski, W. (2026). From testbeds to high-stakes work: A review of human–AI teaming domains and teaming factors. *Frontiers in Robotics and AI*, 13, 1733942. <https://doi.org/10.3389/frobt.2026.1733942>
- Research Square. (2024). AI risk-management control in human-in-the-loop systems. <https://www.researchsquare.com/article/rs-11227072/v1>
- Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding human-centred AI: A review of its defining elements and a research agenda. *Behaviour & Information Technology*. <https://doi.org/10.1080/0144929X.2024.2448719>
- Sommerville, I. (2015). *Software engineering* (10th ed.). Pearson.
- Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the Longwall method of coal-getting. *Human Relations*, 4(1), 3–38. <https://doi.org/10.1177/001872675100400101>
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*. <https://doi.org/10.1038/s41562-024-02024-1>
- Xu, W., & Gao, Z. (2026). An intelligent sociotechnical systems (iSTS) framework: Enabling a hierarchical human-centered AI (hHCAI) approach. *IEEE Transactions on Technology and Society*.