In the Shadow of Judgment: Mapping Out the Landscape of Human–AI Decision-Making Through a Systematic Review

Yixin Bai, Taehyun Yang, Zhongzheng Xu, Dayeon Ki, Ziyi Wang, Yu Hou, Fumeng Yang
{yixbai01,taeyang,zxu169,dayeonki,zoewang,houyu,fy}@umd.edu
University of Maryland, College Park

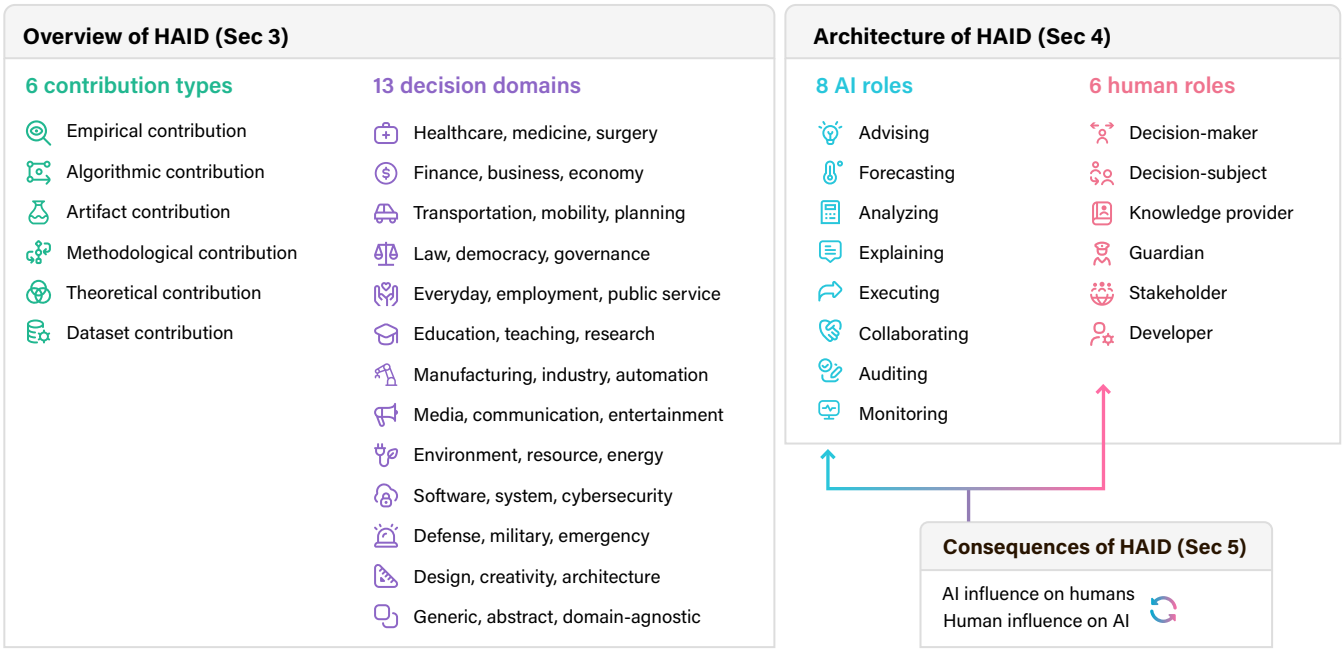


Figure 1: Our systematic review of human–AI decision-making (HAID) literature maps out 6 contribution types, 13 domains of application, and 8 AI and 6 human roles, along with how these roles influence one another.

Abstract

Research on human–AI decision-making (HAID) has grown rapidly across domains. However, the literature on this topic remains fragmented, and conceptual ambiguities hinder systematic synthesis and cross-domain comparison. We present a systematic review of HAID publications from 2016 to 2024. Starting with 269,060 records collected from 20 publishers and databases, we conducted 5 rounds of relevance and quality screening, and ultimately analyzed 1,493 articles in breadth and depth. We contribute: (1) a systematic categorization of the literature, including 6 contribution types, 13 domains of application, and 4 decision levels; (2) taxonomies of 8 AI roles and 6 human roles, their inputs to decision-making, and their influences on each other; and (3) the identification of 4 research gaps and pathways forward. Our efforts unify scattered work into a coherent framework, strengthen the descriptive and rhetorical power of this field, and lay the groundwork for future research and policy development.

Keywords

Decision-making, Human-AI Interaction

1 Introduction

The development of artificial intelligence (AI) has been expanding for over half a century, and in recent years, these AI systems have begun to transform our society. At the heart of this transformation is their power to *shape*—and in some cases *determine*—**decisions** across levels of society, from individuals to organizations and institutions. The stakes extend far beyond individual outcomes: which patients receive treatments [63], which students gain educational opportunities [309], which loan applicants receive approval [311, 565], and which policies governments prioritize [613].

However, AI systems are often prone to errors, and these mistakes can have severe consequences such as misdiagnoses and

This is an unpublished manuscript. It was submitted to CHI 2026 and went through a Revise & Resubmit cycle, but was ultimately not accepted. You are welcome to cite this work as follows.
@article{bai2026haid,
title={In the Shadow of Judgment: Mapping Out the Landscape of Human–AI Decision-Making Through a Systematic Review},
author={Bai, Yixin and Yang, Taehyun and Xu, Zhongzheng and Ki, Dayeon and Wang, Ziyi and Hou, Yu and Yang, Fumeng},
journal={Manuscript},
year={2026},
doi={https://doi.org/10.36227/techrxiv.176963824.45099699/v1}
}

wrongful arrests. Equally important is to ensure that AI systems serve human interests rather than merely optimize metrics. These dual concerns make fully automated decision-making neither entirely feasible nor desirable. As a result, humans and AI systems increasingly operate together to reach decisions, a paradigm we understand as *human–AI decision-making* (HAID).

With a growing body of literature on HAID, considerable ambiguity persists about its fundamental nature and scope. If HAID is to be understood as an impactful paradigm across domains, we first ask: **which domains study HAID, and what kinds of contributions have emerged?** In particular, what decision tasks are examined, and how do emphases vary across domains? As we map this landscape across domains and decision tasks, we confront a more fundamental question: **what, at its core, constitutes HAID?** What roles do humans and AI play as constitutive elements of the process, and how do they interact to shape decisions? Finally, a further question arises: **what consequences follow from HAID?** How do these human–AI interactions affect outcomes and influence the broader systems in which decisions are embedded? Addressing these questions is not merely a matter of definition but also equipping researchers and practitioners with the conceptual clarity needed to advance the field in meaningful directions.

Seen in this light, we conduct a systematic review of HAID publications from 2016 to 2024, extending beyond conventional HCI conferences to capture the scope of the literature. We surveyed 20 publishers and databases, spanning 3 computing publishers (e.g., ACM), 7 major AI venues (e.g., NeurIPS), 7 general publishers (e.g., Elsevier), and 3 disciplinary publishers (e.g., APA), and gathered 269,060 initial records. We performed 5 rounds of systematic screening and filtering to ensure quality and relevance, finalizing the corpus at 1,493 papers.

To capture the field in breadth, we first depict an overview of contributions, domains, and decision tasks. We extend the taxonomy of HCI contributions [584] to catalog the *contribution types*, revealing the dominance of empirical and algorithmic works (Sec. 3.1). We identify 13 *domains* where HAID has been applied and classify decisions in these domains by their social scope—from individual choices to institutional policies (Sec. 3.2).

As a deeper inquiry into what constitutes HAID, we distill the roles that humans and AI systems serve in a decision-making process (Sec. 4). We identify *8 AI roles* and *6 human roles* that characterize HAID research across domains, extracting *their inputs* to decision-making and mapping them onto the decision-making process model [354]. While AI roles are concentrated in the core steps of decision-making, human roles extend to the broader HAID ecosystem, including oversight, governance, and stakeholder influences that collectively shape the decision-making process. To understand the consequences of HAID, we examine the *influences* that humans and AI have on each other in decision-making, revealing trends of co-evolution (Sec. 5). Drawing on these observations, we offer four perspectives on gaps in the literature and pathways to address them (Sec. 7).

Contributions. We map out a landscape of HAID that outlines both the scope of this field and the current state of knowledge. In doing so, we contribute in three ways. We provide ❶ a systematic categorization of the literature, bringing together previously scattered work into a unified framework and offering readers a coherent understanding. We develop ❷ a shared vocabulary and set of concepts that describe HAID processes, strengthening the field's *descriptive power*—the ability to study, characterize, and report on HAID, as well as its *rhetorical power*—the capacity for the research community to communicate and compare key elements across studies. We identify ❸ key gaps in the literature and provide actionable recommendations. While our work is primarily descriptive, it lays the foundation for future theoretical, empirical, and methodological contributions. Given the impact of HAID across society, this systematic understanding has implications for both research advancement and policy development. To support future work, we release our corpus, source code, annotation tool, and codebook at https://doi.org/10.17605/OSF.IO/U5QJH and deploy an interactive web interface at https://fig-x.github.io/haid/.

1.1 Definitions

To align readers from different backgrounds, we first define human–AI decision-making. This requires articulating three components: what decision-making is, what qualifies as AI, and how humans and AI combine in decision-making.

Decision-making is "the cognitive process of choosing between two or more alternatives, ranging from the relatively clear-cut to the complex." [14, 415] For a process to qualify as decision-making, alternatives must be presented, and a decision-maker must select one and act on it. This is related to, but different from, *judgment*, which involves forming an assessment without necessarily choosing among alternatives [139, 415]. For example, a prediction is not itself a decision, but it may constitute a judgment or be considered an alternative. Decision-making processes involve multiple stages, and we adopt the widely used Lunenburg's model [354] with 6 detailed steps: *identifying the problem*, *generating alternatives*, *evaluating alternatives*, *choosing an alternative*, *implementing the decision*, and *evaluating decision effectiveness*.

The concept of **AI** resists singular definition [231, 581, 598]. The literature offers many definitions [231]. Here we understand AI as a collection of technologies [598] that "replicate human capabilities" to varying degrees [581] and "interact with us and act on our environment, either directly or indirectly" [231].

We synthesize these components by drawing from collaborative decision-making, which "aggregates, rather than compromises, the understandings of decision makers" [417]. We define **HAID** as a decision-making process where both human and AI contribute inputs, whether those contributions occur synchronously or asynchronously, directly or indirectly. This definition emerged from screening and analyzing thousands of articles, reflecting our understanding of the boundary between what constitutes HAID and what does not.

2 Methodology

To begin, we first describe our selection and coding processes. We adhere to the Statement on Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020 Statement) [419] and start with searching, filtering, and screening, then proceed to qualitative coding. We show an overview of our process in Fig. 2.

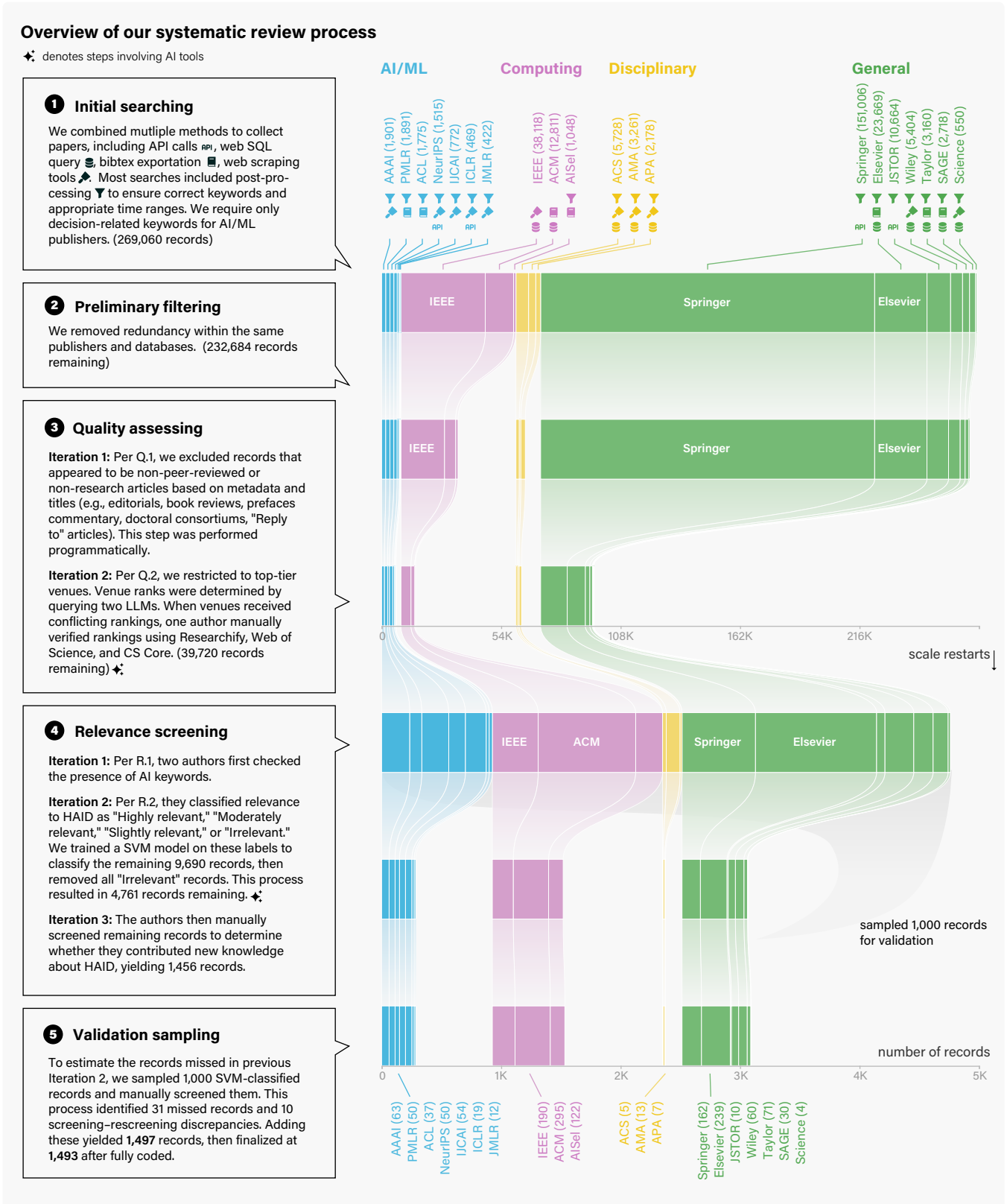


Figure 2: Overview of our systematic review process: We begin with broad searches and then narrow to top-tier venues. Through a combination of manual screening and ML-based classification, we reduce the collection to 4,761 records. After manual screening and qualitative coding, we finalized a corpus of 1,493 records.

2.1 Inclusion and exclusion criteria

We set criteria along two dimensions—*relevance* and *quality*—to guide our screening and selection processes. First, to ensure **relevance**, a paper must meet the following:

- R.1 The title and abstract must contain both AI- and decision-related terms. When an abstract is absent, we conduct a preliminary full-text review to confirm their presence.
- R.2 The paper must involve both humans and AI in a decision-making process (i.e., meets HAID definition). We exclude works that focus on predictions and modeling without a clear connection to decision-making.

Also, to maintain **quality**, we require:

- Q.1 The paper must be peer-reviewed and be published in an academic journal or conference, with a minimum length of 3 pages. While acknowledging exceptions, this threshold ensures sufficient details and substantive contributions for coding purposes.
- Q.2 The paper must be published in a top-tier venue in its domain. Although venue rank is only a proxy for quality, this helps ensure the rigor and credibility of selected papers and preserves a manageable corpus size.

2.2 Initial searching and preliminary filtering

We included a range of publishers and databases, moving beyond conventional HCI venues to capture a broader body of literature. Our R.2 criterion above guarantees the HCI relevance of the resulting papers.

Publisher and database. We surveyed 20 publishers and databases, and grouped them into four categories. We began with 3 computing databases: ACM Digital Library, AIS eLibrary, and IEEE Xplore, which cover the majority of human-centered computing research. We also collected from 7 major AI/ML venues: NeurIPS, ICLR, PMLR, JMLR, ACL Anthology, IJCAI, and AAAI[1]; following Q.2, we excluded tracks like high school projects at NeurIPS and retained only recurring conferences and symposia at AAAI. We searched 7 major interdisciplinary and general publishers: Springer Nature, Elsevier (ScienceDirect), JSTOR, Wiley, Taylor & Francis, SAGE Journals, and Science. Finally, we queried 3 prestigious disciplinary databases: ACS, AMA, and APA[2] to have coverage of domain-specific research. Full details on the inclusion of specific conferences and tracks are provided in supplementary materials.

Scope. Two authors brainstormed a set of keywords for AI and decision-making. Following R.1, we required a record to contain at least one AI keyword and one decision keyword (case insensitive) in its title or abstract. As such, a typical search string was: `(Artificial Intelligence or Machine Learning or Deep Learning or Natural Language Processing or Reinforcement Learning or Large Language Model or Large Language Models or AI or ML or DL or NLP or LLM or LLMs) and (decide or decision or decisions)`. We included plural forms because several databases were sensitive to them. However, we require only decision keywords for AI/ML venues, since papers in these venues typically employ more specific or emerging AI terms rather than generic ones; plus, we assume reviewers at these venues already ensure the relevance of accepted papers to AI. Finally, we restricted publication dates to 2016–2024 for two reasons: a prior survey by Lai et al. shows that HAID research began to grow around 2016 [300], and most of our searches were conducted between November and December 2024, with a small number in January 2025.

Process. We combined multiple means to obtain records from these publishers and databases. Many publishers and databases did not support direct exportation of search results, and returned records often exceeded maximum thresholds (e.g., ACM returns only the first 2,000 records); some imposed constraints on the number of keywords or supported fields (e.g., Elsevier allows only 8 keywords). We performed searches in small batches and employed web scraping where necessary. We retrieved all recent NeurIPS and ICLR via the OpenReview.net APIs and had to search in full text with the restrictions on Springer Nature API. Due to these inconsistencies, we applied post-processing to filter by keywords and to verify publication dates. In total, we obtained **269,060** records, and Fig. 2 shows a breakdown.

Preliminary filtering. We programmatically removed duplicates within each publisher or database (matching title and author strings), reducing the corpus to **232,684** records.

2.3 Quality assessing

To ensure the **quality** of the resulting paper, we conducted two rounds of filtering.

Iteration 1 (Q.1). We filtered out records that appeared to be non-peer-reviewed or non-research articles based on metadata and titles. This step removed editorials, book reviews, prefaces, commentary, doctoral consortia, corrections, titles beginning with "Reply to: ", among others. We also excluded records under 3 pages where page metadata appeared reliable.

Iteration 2 (Q.2). We restricted records to top-tier venues. We encountered 19,498 venue names with considerable variation.[3] Given this scale, we employed LLMs for initial assessment and validated this approach with human judgment. We prompted both GPT-4o and Claude 3.5 to categorize each name variant into top-, mid-, and low-tier. When the two models disagreed, one author provided the third vote using three ranking systems: CORE Conference Ranking [1], Researchify [2], and Web of Science [3], and a top-tier venue must be ranked as Quartile 1, A-, or above by at least one source; unlisted venues were evaluated by comparing impact factors. We retained venues with two top-tier votes.

[1] Conference on Neural Information Processing Systems (NeurIPS), International Conference on Learning Representations (ICLR), Proceedings of Machine Learning Research (PMLR), Journal of Machine Learning Research (JMLR), Association for Computational Linguistics (ACL), International Joint Conferences on Artificial Intelligence (IJCAI), and Association for the Advancement of Artificial Intelligence (AAAI), respectively. Note that PMLR includes conferences such as AISTATS and ICML, and ACL Anthology includes several venues such as ACL, EMNLP, NAACL, COLING, and CoNLL.

[2] American Chemical Society (ACS), American Medical Association (AMA), and American Psychological Association (APA).

[3] For example, "CHI Conference on Human Factors in Computing Systems," and "Conference on Human Factors in Computing Systems."

Validation. To validate our use of LLMs in Iteration 2, we randomly sampled 100 venues from each tier (300 in total) and applied the same human-judgment method. We found LLMs were generally conservative: 0% of LLM low-tier and 32% of mid-tier venues were upgraded to top-tier upon human judgment. Only 3% of LLM top-tier venues were downgraded. The resulting venues maintain broad coverage across domains. Removing 23 duplicates, we arrived at **39,697** records. The prompts are provided in Appx. F.

2.4 Relevance screening

To ensure the **relevance** of the resulting papers, we combined manual screening with machine learning methods.[4]

Iteration 1 (R.1). We verified the presence of desired keywords, as previous programmatic filtering had captured irrelevant matches (e.g., "ai" in gain). This iteration led to 24,430 records.

Iteration 2 (R.2). Two lead authors classified records into *Highly relevant*, *Moderately relevant*, *Slightly relevant*, or *Irrelevant* based on titles and abstracts, with one author reviewing a subset. They followed the same protocol, met weekly to discuss uncertain cases, and screened 14,537 records in a span of 10 weeks. Multiple factors then motivated follow-up automated screening: (1) the remaining records (mostly Elsevier, Springer Nature, and IEEE) contained massive irrelevant and marginally relevant papers that added little new insight. (2) Borderline cases remained ambiguous after extensive discussion. (3) Fatigue threatens reliability, and automation helps preserve it by directing human judgment where it matters most. (4) The best-performing classification model we explored, a linear SVM, achieved 83.86% soft accuracy (all *Relevant* vs. *Irrelevant*) and 80.80% overall accuracy in cross-validation. Finding the accuracy satisfactory, we applied the model trained on all screened records to the remaining 9,690 records, retaining only those predicted as relevant (i.e., not *Irrelevant*). The authors then screened the other 203 records lacking abstracts via full-text. This step yielded 4,971 records.

Iteration 3 (R.2). We subsequently found many remaining abstracts were insufficient for providing knowledge about HAID, merely using decision-making as a motivation or implication. The same two authors reassessed each abstract using a guiding question: "Could this work advance our understanding of HAID?" This requires a paper focusing on decision-making with both human and AI inputs (the HAID definition in Sec. 1.1). We operationalized this by screening a few hundred papers separately, discussing weekly to converge understanding, and compiling exclusion reasons (see Appx. E). Common reasons include: (1) lack of human elements, (2) no clear connection between AI and human inputs, or (3) not focus on decision-making. Occasionally, abstracts were removed for violating other criteria (e.g., one page). They then screened remaining titles and abstracts separately, discussed uncertain cases, and cross-checked each other's screening results. This yielded **1,456** records after excluding survey and opinion pieces.

Validation. We stratified a sample of 1,000 papers by database and publisher from the 9,690 SVM-classified records. This subset enabled us to (1) estimate missed papers and (2) assess screening reliability. Of these, 863 had been classified as *Irrelevant* and previously removed after Iteration 2. The same two authors manually screened them using the Iteration 3 criteria (i.e., single-stage screening), identifying **31** missing records. The same qualitative analysis in Sec. 2.5 was applied, but revealed no new codes, indicating content saturation. We analyzed the effects on corpus and code distributions in Appx. C. In later sections, we interpret results and draw conclusions with these uncertainties in mind. Also, the authors rescreened the other 137 papers[5] and found **17** discrepancies: 10 from irrelevant to relevant and 7 the reverse. Because Krippendorff's α and similar metrics are unreliable with skewed distributions like ours [445, 556], we also report Gwet's AC2 [445] and percentage agreement (PA): 0.73 (α), 0.77 (AC2), and 0.88 (PA). These indicate **good** screening reliability, particularly given that these sample papers were already marginal and ambiguous in relevance. We discuss SVM errors and feature importance in Appx. A and B. Adding 41 recovered papers, we arrived at **1,497** records.

2.5 Qualitative coding

General procedure. The two lead authors agreed on the dimensions to extract: (1) *contribution types*, (2) *domains*, (3–4) *human and AI roles*, (5) decision levels, (6–7) *humans' and AI's influences on each other*, and (8) *decision environment*. The first four capture the **breadth** of literature: *contribution types* and *domains* define a research paper, while *human and AI roles* define HAID. These appear in all papers and are typically summarized in titles or abstracts, so we coded them primarily from abstracts, consulting full texts when necessary. In contrast, *humans' influence* and *AI's influence* reflect the **depth** of literature. These dimensions require full-text analysis and appear primarily in papers involving human subjects, controlled experiments, novel systems, or theoretical frameworks. The two authors identified these papers (N = 483) by screening independently and cross-checking their results (see Sec. 2.4; Krippendorff's α = 0.66, Gwet's AC2 = 0.93, PA = 0.94 in validation). These papers also provided examples for *humans' and AI's inputs* and *decision environments*, which we treated as illustrative rather than systematically coded. Finally, *decision levels* reflect both **breadth** and **depth** and were coded from both abstracts and full texts.

General process. Two primary coders (the two lead authors) and four secondary coders were involved. Each primary coder reviewed a subset; they together developed lower-level codes, merged them into higher-level categories, and finalized a codebook with non-mutually-exclusive labels. Primary coders then provided initial codes for all papers. Secondary coders refined these initial codes and discussed with primary coders to address disagreements.

LLM assistance & validation. We decided on using LLM assistance in coding and also validated this approach. We prompted GPT-4o and Claude 4.5 (3.5 was deprecated) with the codebook, ran each three times, and aggregated outputs as votes to mitigate uncertainty (e.g., decision-maker (3)). To have a reference (our "gold standard"), one secondary coder independently coded a 10% sample, and reconciled with primary coders. Three other secondary coders used different LLM-assisted approaches for the same subset, adjusting primary coders' initial codes using GPT results, Claude results, or both. Because multi-label reliability calculation remains an open

[4]Please also read our reflection on the method in Sec. 7.6.

[5]The screening and rescreening processes were months apart.

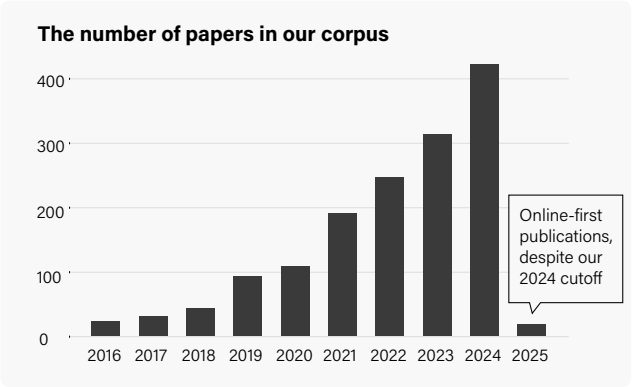


Figure 3: Distribution of the papers in our corpus, showing an increasing trend recently.

problem [363], we computed and averaged per-label scores for each dimension. Compared with the reference, providing both LLMs' outputs yielded the highest reliability (Krippendorff's α = 0.26–0.80, Gwet's AC2 = 0.63–0.97, PA = 0.77–0.98), and was applied to the remaining corpus by two secondary coders, each analyzing a subset. One dimension and 8 labels with AC2 < 0.66 received additional instructions along with noted LLM biases. Primary coders then refined and finalized codes through asynchronous discussion. Reliability scores between initial, adjusted, and final codes were 0.64–0.88 (α), 0.83–0.98 (AC2), and 0.89–0.99 (PA). All scores, prompts, and scripts are available in Appx. D, F, and supplementary materials.

Final corpus. Excluding 4 survey/opinion pieces identified from coding, we finalized at **1,493** records.

3 Overview of HAID

We depict a distribution of our corpus in Fig. 3. As anticipated, the number of relevant publications is growing at an accelerated rate, demonstrating the timeliness of a systematic review. We begin by showing *contribution types* and *domains of decision-making*, along with the corresponding counts. As HAID research continues to evolve, mapping contributions and domains provides an overview of existing knowledge. Given our corpus size, we cite only example papers or those with at least 100 Google Scholar cites by the end of November 2025.

3.1 Contribution types

We adopt Wobbrock and Kientz's taxonomy of 7 research contribution types in HCI [584], extending it with an additional category, *algorithmic contribution*, to more precisely describe the contributions from AI/ML venues. We assign one primary type unless two types are genuinely equivalent. Excluding survey and opinion papers, we present 6 types below.

At a glance

Empirical and *algorithmic* works dominate across AI/ML, computing, and general publishers, while *methodological*, *artifact*, and *theoretical* works provide bridges between them. Nevertheless, *datasets* remain rare.

Empirical contribution (N = 555). As the largest portion in our corpus, many works make empirical contributions through controlled human-subject experiments [165, 253, 344, 358, 476, 484, 513, 601], questionnaires [78, 108, 185], interview studies [15, 284, 607], observational research [158, 364], algorithmic and model evaluations [467, 484], case studies [120], and mixed methods research [499]. They provide empirical insights into factors, methods, and strategies that shape HAID (e.g., explanation [71, 211, 479]), alignment between human and model performance [45, 266, 582], and the social and organizational contexts [96, 504, 505]. This type was typically the second largest prior to 2022, and has since surpassed *algorithmic contribution*, suggesting increased interest in recent years.

Algorithmic contribution (404). A substantial body of work contributes new AI/ML algorithms, models, or computational approaches. These works often explore algorithmic pathways to assist humans in decision-making, including generating model interpretation [27, 593], finding solutions for ethical and fairness constraints [73, 168, 346, 570], and making domain specific predictions [97, 482, 488, 628]. While computational efficiency and performance are the primary focus, some works evaluate and formalize human experience within decision loops [510, 635].

Artifact contribution (257). These works contribute implemented systems, tools, interfaces, and design solutions. Decision Support Systems (DSSs) are most common, integrating expert knowledge for clinical diagnosis [50, 173, 188, 423, 603], treatment selection [312, 413, 447], and policymaking [328]. Beyond DSSs, artifacts manifest as conversational chatbots for IT support [22], visualization tools for counterfactual explanations [187, 647], and immersive systems for basketball decision-making [535]. Most remain domain-specific, with few achieving domain-agnostic applicability [552, 647].

Methodological contribution (232). These works provide new research methods, evaluation frameworks, analytical approaches, and design principles. The topics parallel those of *algorithmic contributions*, including interpretability and explainability frameworks [279, 351], fairness-aware methods [389, 444], and human-in-the-loop frameworks [8, 618], yet emphasize generalizability across contexts. Some propose fairness frameworks [117, 624] or auditing methodologies [13], while others develop evaluation methods within decision-making contexts [254, 303, 392, 401]. Domain-specific methodologies also emerge, such as design guidelines for clinical decision support [221].

Theoretical contribution (128). Theoretical works articulate the conceptual, normative, and philosophical foundations of HAID. Many formalize core concepts (e.g., fairness [105, 296], explanation [560], and accountability [94, 102]) through causal models [92, 406], information-theoretic measures [310], and legal theory [291], advocating safeguards such as contestability and reviewability [112, 245, 356, 406, 613]. Others develop ethical frameworks and models of human–AI complementarity [33, 132, 138, 289, 335, 345] or examine computational limits in capturing human judgment, particularly indecision and hard choices [190, 289, 375].

Dataset contribution (13). Decision benchmarks are rare, often paired with *empirical* and *algorithmic contributions*. Most target domain-specific decisions (e.g., clinical practice [209, 238, 337], driving [508], and game play [313]), though some others extend to embodied environments [393]. However, no datasets address interactive or collaborative decision-making.

3.2 Domains of decision-making

We identify 13 application domains where HAID has been explored (see Fig. 4). For abstracts that do not explicitly state domains, we extract domains from their decision tasks and datasets used. To clarify decision levels, we adopt a four-level taxonomy from decision-making literature: **individual** decisions made by a single person for themselves [68], **operational** decisions embedded in day-to-day routines and practices [532], **organizational** decisions made within or on behalf of organizations [68], and **institutional** decisions that establish standards or norms [379]. Although drawn from different traditions, this hierarchy provides a consistent vocabulary.

Each paper can span multiple domains. For example, a fairness method may apply to both finance and employment, while a work on medical policy falls under both healthcare and law. We present *institutional* decisions within each domain, but also discuss law as a domain when it is the primary focus.

At a glance

The distribution across domains reflects AI's broad penetration into decision-making contexts, with medical applications comprising the largest share. *Operational* decisions dominate overall, though their prevalence varies across domains. High-stakes domains (e.g., defense, law) tend to frame HAID around accountability and risk, while everyday contexts (e.g., media, education) position it as a supportive tool. Fairness and governance concerns surface across domains, with exceptions in defense and design, where the nature of their decision tasks makes such framing less relevant. Finally, the prevalence of generic works suggests either efforts toward generalizable frameworks or a lack of ecological validity in the literature.

Healthcare, medicine, surgery (N = 480). Healthcare is the most extensively represented domain, with nearly one-third of papers addressing medical decision-making. *Individual* decisions include personal diabetes management [400, 540]. Clinical *operational* decisions predominate: clinical diagnosis [37, 64, 80, 86, 137, 343, 361], surgery prioritization [605], selecting treatment [304, 399], emergency department triage [308], drug dosage optimization [407], ICU admission [606], and ventilator configuration [286]. *Organizational* and *institutional* decisions include emergency department management [150] and COVID vaccine allocation [213]. Decision tasks may become cross-level in certain circumstances: admitting a patient can shift from *operational* to *organizational* or *institutional* levels when resource scarcity or accountability is at stake [606].

Finance, business, economy (193). Financial decision-making often appears in managerial or business literature. *Individual* decisions include accepting AI shopping advice [77, 87, 99, 366, 367, 380], or peer-to-peer lending or investing [126, 131]. *Operational* decisions involve approving loans through credit scoring [174, 217, 240, 418], making sale strategy [46], setting dynamic prices in retail [323, 372, 381], rebalancing fund portfolios [430, 500, 640], and deciding trading strategies [131, 290, 468, 625]. *Organizational* decisions involve allocating venture funding [373, 402, 641], and team changes in management [297]. At the *institutional* level, research addresses governance frameworks and legitimacy of algorithmic decisions in markets [47, 370, 514].

Transportation, mobility, planning (189). Many works focus on self-driving cars, examining decisions from individual vehicle behavior to systemic transportation. *Individual* decisions include lane-changing, braking, ramp merging, and intersection navigation [162, 220, 326, 391, 563, 590], with attention to human-like reasoning [241, 342, 347], uncertainty [229, 432, 528], and ethical dilemmas in crash or trolley scenarios [30, 168, 169, 292, 543]. *Operational* decisions involve coordinating connected fleets for ride-sharing and traffic [210, 270] and optimizing routing and charging for electric vehicles [357, 586, 643]. *Organizational* decisions address aviation operations [66, 127], cargo allocation [134, 453], and maritime navigation [52, 551, 557]. *Institutional* decisions extend to infrastructure maintenance and resource allocation [438, 577], as well as traffic law and public trust for autonomous vehicles [336, 339, 604].

Law, democracy, governance (122). Legal contexts rarely involve purely *individual* decisions; most decisions are *operational*. Examples include criminal sentencing and parole decisions informed by AI bail and recidivism risk assessments [25, 135, 307, 334]. *Organizational* decisions include policy development through impact analysis [285], fraud detection in unemployment programs [177], and AI adoption in public sector [265]. At the *institutional* level, work addresses public trust in AI for election monitoring [263, 287, 525], and AI governance spanning all domains.

Everyday, employment, public service (116). This domain spans from personal choices to social care and public service, often co-appearing with finance and law. *Individual* decisions include selecting fashion, movies, restaurants, or tasks on crowdsourcing platforms based on algorithmic recommendations [140, 212, 236, 646]. *Operational* decisions dominate, notably in using AI for candidate hiring [16, 172, 230, 435, 483] and college admission [141, 309];[6] these also include welfare decisions for child protection [259, 469]. *Organizational* decisions involve workforce management in companies [455], such as identifying high-potential employees [387]. Finally, at the *institutional* level, AI adoption in public services raises debates around fairness and accountability [262, 550].

Education, teaching, research (74). Students, teachers, and researchers increasingly use AI to support decision-making. *Individual* students select universities and programs with AI assistance [100]. *Operational* decisions include teachers utilizing AI to judge student performance [107], researchers using AI to guide methodological choices in qualitative analysis [166, 416], and reviewers using AI

[6] Considerations in college admission are similar to hiring, so we place it here rather than under education.

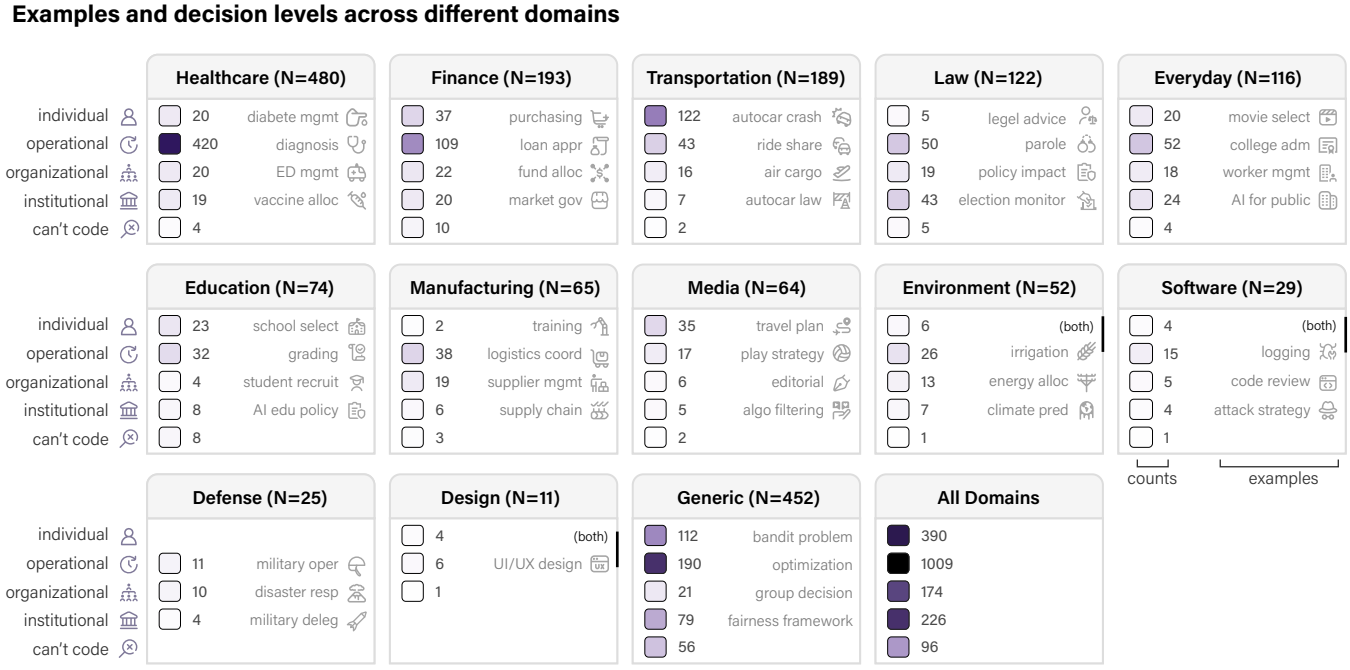


Figure 4: Examples and decision levels across application domains: *operational* decision-making dominates, while *individual* decision-making is prevalent in transportation and media.

for peer review decisions [19, 136, 293, 294, 524]. *Organizational* decisions include universities employing AI to recruit students [100]. *Institutional* work addresses education policy and governance, raising broader questions about fairness, discrimination, teacher roles, and student vulnerability when using AI in decision-making [116, 295, 441, 442].

Manufacturing, industry, automation (65). This domain covers robotics and supply chains. *Individual* decisions include workers navigating their reliance on automation during training [431]. *Operational* decisions include production scheduling [591], logistics coordination [318], and maintenance strategy [18]. Similarly, *organizational* decisions concern using AI to manage suppliers [28, 443] and risks [130, 446], as well as adopting AI [119, 144]. *Institutional* work explores AI-driven digital twins [203, 369], as well as fairness and accountability in manufacturing [70].

Media, communication, entertainment (64). *Individuals* can assess news credibility with AI support [454], make travel and streaming choices with AI recommendations [189, 276], and make moves in games and sports (e.g., chess) with AI advice [180, 376, 390, 507, 529, 566]. *Operational* decisions include content moderation through AI identifying harmful posts [62, 75, 299] and coaching strategies informed by AI analysis [84]. *Organizational* decisions include editorial choices about AI in publishing [503] and game management for player retention, matchmaking, and server optimization [121, 588]. *Institutional* work explores regulatory frameworks for social media filtering [243] and LLM interventions against misinformation [164].

Environment, resource, energy (52). This domain spans local farming to global sustainability. At the *individual* and *operational* levels, farmers use AI to decide on irrigation, harvest timing, and crop treatment [109, 269, 396, 458], as well as for soil management [10, 378]. *Operational* decisions also include utility providers using algorithms for load shedding, repairs, and resource deployment during disasters [152, 234, 644]. *Organizational* decisions include energy providers using AI for demand allocation, market negotiation, and maintenance prioritization [306, 638, 639]. *Institutional* decisions include policymakers employing AI to balance food–energy–water trade-offs, interpret climate forecasts, and guide renewable energy adoption [191, 541, 575, 623].

Software, system, cybersecurity (29). Decision tasks in this domain range from everyday software engineering to strategic cybersecurity. *Individual* developers use AI for micro-decisions such as logging statement placement [330] and library selection [340]. *Operational* and *organizational* decisions include code review [630], software architecture design [515], and IT task management [22]. *Institutional* decisions include cybersecurity organizations using AI to reallocate alerts and develop defensive strategies [489].

Defense, military, emergence (25). In this high-stakes domain, decisions are rarely individual. *Operational* decisions include military operators using AI to interpret physiological data [459], detect abnormal behavior in surveillance [128, 129], and navigate hazardous situations [393]. *Organizational* decisions include emergency facility planning and resource deployment [420, 523], as well as integrating AI into emergency response systems [385, 492, 595, 631]. *Institutional* decisions extend to strategic military choices, such as air defense planning with AI support [633] and delegating military operations to AI [255, 452].

Design, creativity, architecture (11). *Individual* designers use AI as a creative support tool for UI/UX design [371], product development [509], and energy-efficient building design [83, 574] in their *daily work*. Few works address *organizational* or *institutional* decisions.

Generic, abstract, domain-agnostic (452). This category covers works that do not articulate a specific application area, addressing decision-making in principle rather than practice. For example, bandit problems and general image classification approximate *individual* decisions [57, 422, 481, 530]. Domain-agnostic algorithms for scheduling or resource optimization assist in *operational* decisions [133]. *Organizational* research explores group decision-making theories and frameworks [118, 491, 498]. Finally, many discuss fairness, governance, or ethical principles in algorithmic decision-making meant to apply across domains [111, 194, 298, 329, 394, 610, 622].

4 Architecture of HAID

Answering what constitutes HAID requires moving beyond domains and tasks, and we approach this from the perspective of what functions human and AI roles serve in the decision-making process. We categorize *AI roles*, *human roles*, and distill their respective *inputs* to decision-making as the core elements of HAID. These elements specify who makes decisions, what information is provided in the process, and where the boundaries between humans and AI are drawn. To situate these roles within decision-making itself, we map them onto the 6-step decision-making process model [354] (see Fig. 5). Beyond the process itself, we also consider the broader *environments* in which decisions operate. Together, these components form an architecture for understanding how humans and AI operate in decision-making.

4.1 AI roles & inputs

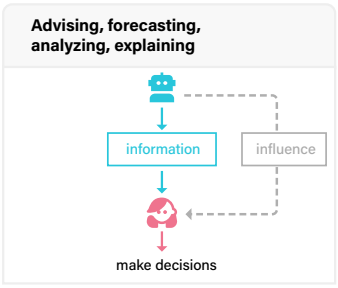
We identify 8 AI roles from our corpus: *advising*, *forecasting*, *analyzing*, *explaining*, *executing*, *collaborating*, *auditing*, and *monitoring*. We group them into 3 categories based on their primary functions: ***information-oriented roles*** that provide information, ***action-oriented roles*** that perform actions, and ***evaluation-oriented roles*** that provide assessments. Individual works may involve multiple roles (e.g., using multiple AI models or including several experiments).

At a glance

While *advising*, *forecasting*, and *executing* roles dominate, few works support all stages of decision-making. Support for the final stage—and *evaluation*—is particularly sparse. Different roles often overlap, and much literature treats AI as a monolithic entity that may aid in unspecified stages of decision-making. This vagueness makes evaluation both difficult and ambiguous: a tool assessed as "decision support" (implying *advising*) may provide only forecasts, appearing ineffective when measured against misplaced expectations. While broader institutional mechanisms exist (e.g., policy), *individual* and *operational* decision-makers often lack support to understand the consequences of their choices.

Information-oriented roles:
These roles provide information on decision alternatives, differing in information type and contribution to the final choice, while humans ultimately retain responsibility for the final decision.

Advising (N = 1,000). An *advising* role gives strategic guidance for generating and evaluating alternatives. This role assumes that AI can access broader information, offer domain-specific or complementary expertise (e.g., models trained on hundreds of cases vs. clinicians with limited experience [155, 319, 617]), or reduce certain human bias [193, 268, 341, 434]. Advising often co-occurs with *forecasting* [5, 301, 484, 519, 578], *analyzing* [63, 125, 160, 225, 607], or *explaining* [27, 104, 153, 252, 273, 424, 449, 463, 475, 558] when framed as predictions, structured analyses, or are accompanied by explanations.



information-oriented

AI input. This role provides prescriptive input for decision alternatives (e.g., suggestions, options, or preferences). Common forms include recommendations [180, 485, 617] in medicine [155, 186, 308], finance [21, 157], transportation [368], and employment [434]. Advisory inputs may also appear as suggestions in data labeling [123] or as second opinions for comparison with human judgment [348]. They can incorporate delegation or deferral mechanisms assigning responsibility to humans [367, 431], or tailor to individual preferences or specific contexts [76, 362, 368].

Forecasting (526). Unlike prescriptive roles, *forecasting* provides expectations about future states without necessarily connecting to specific decision alternatives. Forecasting frequently accompanies *advising* or *executing* roles [65, 95, 464, 611]. Some research may not state how forecasts integrate into decision processes [634]. In these instances, forecasts alone may support problem identification by indicating what might happen [354].

AI input. This role inputs predictive information about future outcomes or states without prescribing an action. Common forms include classifications, such as detecting tumors from skin lesions [208], risk estimations [6], and predictions of treatment outcomes or disease trajectories [382, 564, 600]. Uncertainty emerges as a recurring theme in forecasting roles [198, 439], which we discuss as a cross-role input at the end of this subsection.

Analyzing (465). An *analyzing* role transforms complex information into digestible formats. It can serve most decision-making stages by providing information about problems, alternatives, and outcomes. Examples include synthesizing documents into reports or tables [110, 157, 384], performing data analysis [228, 446], identifying patterns or features (e.g., in medical images [58, 156]), and processing ICU streaming data [386]).

AI input. Analyzing shares similarities with *forecasting* in providing information not directly tied to decision alternatives, but focuses on historical or present states. Inputs can be textual or visual. A representative example is document synthesis that extracts and organizes key points [110, 157, 384, 524]. Examples also

include pattern identifications and feature extractions [275], often presented through visual analytical interfaces to help users detect abnormalities, compare alternatives, or interrogate model outputs [58, 275, 341]. Analyzing can also be embedded in collaborative contexts to support qualitative coding workflows [166].

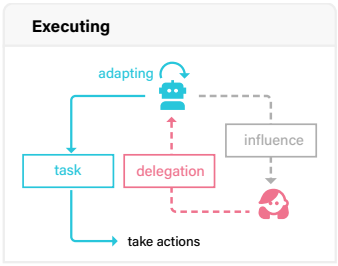
Explaining (341). An *explaining* role produces interpretive information about mechanisms, system behaviors, or rationales to support evaluating alternatives or decision outcomes [354]. Explaining typically does not operate independently but generates explanations for another AI role (e.g., *advising* [55, 398, 569, 626], *forecasting* [235, 457], and *collaborating* [619]). Their explanations may calibrate trust [594], support accountability [642], and enable contestability [516]. For example, saliency maps can highlight features influencing medical diagnoses [267], while counterfactuals can clarify loan denials [72] or reveal model biases [192].

AI input. This role provides justifications or rationales [181] in diverse forms: visual (e.g., feature visualizations [267]), textual [567, 619], and interactive [321, 338]. Explanations may follow feature- or example-based approaches, or both [320, 568]. Other forms show alternative outcomes through counterfactual or what-if explanations [89, 109, 511], or provide contrastive and causal explanations that clarify why one result occurred over another [72, 89, 545, 573]. They can be local or global [398], concise or detailed [143, 597], generic or personalized [44, 478, 480], reflecting different design trade-offs. We direct readers to focused surveys [113, 183, 473] for comprehensive coverage.

Action-oriented roles:

These two roles can take actions and make decisions autonomously or collaboratively, shifting responsibility between humans and AI systems.

Executing (512). An *executing* role replaces humans and directly makes decisions, typically after explicit or implicit delegation from humans [85, 496, 609]. It assumes that AI has or will achieve decisive advantages over humans, producing comparable decisions faster, at larger scales, with fewer resources (e.g., screening job applications [96], operating autonomous vehicles [325], allocating resources [633]), or by delivering better outcomes (e.g., higher accuracy in cancer diagnostics [317]). Such systems often embed human values or approximate human reasoning to maintain alignment with intended objectives [219, 615]. It targets the final decision-making stages—choosing and implementing an alternative [354], thereby shifting responsibility from humans to AI [292].

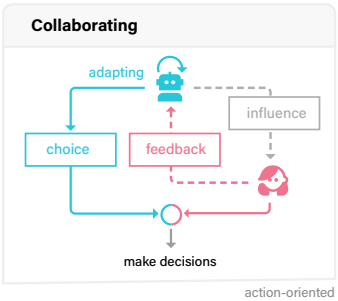


action-oriented

AI input. The input here is the decision or action. The AI system takes available information and directly acts or chooses an alternative [246]. Humans remain responsible for defining objectives, curating data, and reflecting on the consequences of outcomes, aligning with our definition of HAID (see Sec. 1.1).

Collaborating (177). A *collaborating* role works with humans as a partner in dynamic, interactive processes. Unlike roles with one-off communication, this role assumes sustained engagement, where neither human nor AI is sufficient alone [232, 260, 315, 427, 617]. Collaborating can span multiple decision-making stages [397], where humans and AI share agency in choosing and implementing alternatives [281], while jointly learning from outcomes to improve future decisions [239, 472]. Such systems appear as *advising* roles [616], teammates [239, 612], agents/copilots [472, 585], chatbots [527, 554], or hybrid juries [421].

AI input. Collaborating systems provide continuous, adaptive input throughout the decision-making process, adjusting contributions based on ongoing interactions. Common forms include adaptive instruction responding to user behavior [632] and contextual decision suggestions [612]. They may provide systematic cues such as confidence scores [397] and power-sharing signals showing human contributions to decision-making [421]. Increasingly, collaborative inputs use dialogic forms [538] or act as coequal team members in voting scenarios [223, 421, 636], sometimes employing repair strategies (e.g., apologies) to maintain or restore trust [427].

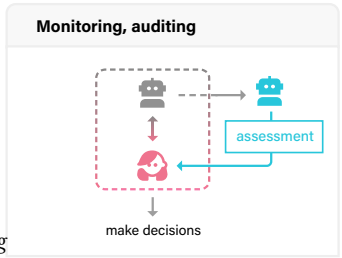


action-oriented

Evaluation-oriented roles:

These roles provide assessment of other AI systems or decision-making processes, differing in the aspects they evaluate.

Auditing (68). An *auditing* role conducts systematic inspection of systems or outcomes against established rules, ethical principles, and fairness standards [116]. It focuses on evaluation against normative criteria, while *monitoring* emphasizes continuous observation. This role aligns most closely with the stage of evaluating decision effectiveness, ensuring that decision outcomes meet broader societal expectations [106, 182, 425]. Rather than providing direct decision support, auditing functions as an accountability mechanism for other roles.



evaluation-oriented

AI input. Auditing provides assessments of decision outcomes and processes to guide future decisions. These inputs manifest as fairness and bias evaluations, including disparity analyses across demographic subgroups [56], and certification signals that indicate compliance with auditing standards [471]. Auditing may employ counterfactual or causal probes to evaluate whether data processing or expert actions satisfy real-world requirements [43, 79].

Monitoring (55). A *monitoring* role continuously observes the states of another system [428] or decision outcomes [176, 307], aligning with the final stage of evaluating decision effectiveness.

AI input. This role delivers ongoing assessment of system performance or decision quality. Inputs can be continuous signals

(e.g., data streams in patient monitoring [352]), or performance metrics evaluating worker decision quality [175]. They can serve as a precursor for *auditing* to evaluate fairness and bias across deployment pipelines [374].

Cross-role input: *Cues and framing*: Systematic and heuristic cues (e.g., disclaimers or identity) can influence decisions [227, 462]. *Uncertainty*: Many works provide explicit textual or visual information about confidence, errors, or risks to help humans weigh alternatives [67, 145, 439]; while being increasingly discussed, we refer readers to focused surveys [29, 40, 572]. *Explanations* also span across *advising* and *executing* roles by mediating human understanding.

4.2 Human roles & inputs

We extract 6 human roles: *decision-maker*, *decision-subject*, *knowledge provider*, *guardian*, *stakeholder*, and *developer*. They situate in the broader HAID ecosystem and may not attach to specific decision-making stages.

At a glance

Humans remain the primary *decision-makers* across all decision-making stages. Their cognitive-affective traits (e.g., attention, biases, trust) shape their understanding of AI outputs and their delegation of autonomy to AI. *Decision-subjects* bear their consequences, typically with limited control over the process. *Knowledge providers* and *developers* train AI with their data and expertise, *guardians* set governance boundaries, *stakeholders* represent broader societal interests; they influence rather than directly participate in decisions.

Decision-makers (N = 1,315). Humans remain the primary decision makers. Often, humans engage with AI to support or guide their choices (e.g., *advising*); in collaborative settings, they may act as one among several decision-makers (e.g., *collaborating*). At times, they delegate authority to AI (e.g., *executing*) and become indirect decision-makers [350]. Across these cases, human decision-makers are assumed to traverse all stages of the decision-making process and to retain at least partial responsibility for outcomes, even when decisions are made by AI.

Input: cognitive-affective traits. Human decision-makers bring cognitive characteristics, emotional states, and individual capacities to bear on decision-making. Implicitly, these traits shape how people attend to, interpret, and decide upon AI outputs—for example, through their attention in the process [397], self-confidence [93], language proficiency [434], prior knowledge and experience [170, 203, 218, 470], intuition [82, 163], strategies [179, 412], economic beliefs [147], hidden biases (e.g., anchoring [411, 448]), or vulnerabilities [274]. Trust is a recurring trait, influencing whether humans rely on, contest, or calibrate their use of AI [23, 280, 589]; we refer readers to focused surveys on trust for comprehensive coverage [215]. These traits can be explicit inputs to AI, as in specifying preferred models [487], or personalizing recommendations [248, 250] to guide responses.

Input: delegation. Humans also determine how much autonomy to grant AI, which can shift across stages and tasks [237, 350]. Delegation may be specific to certain decisions while retaining control over strategic or ideologically charged ones [324], and can be adjusted based on user experience and expertise [216, 274]. However, delegation can be driven by motivations such as responsibility avoidance [161] or the desire to offload unpleasant choices [65].

Decision-subjects (566). Decision-subjects are individuals or groups on whom decisions are made and who are directly affected by decision outcomes. They are often underemphasized in the literature, and foregrounding them is important in order to account for the consequences of decisions. They may be the decision-makers themselves in *individual* contexts. In *operational* and *organizational* contexts, decision-subjects include patients in clinical diagnoses [88, 114, 171, 207, 512, 536, 602, 628], loan or job applicants [174, 240, 418], and defendants or inmates in the justice system [135, 307]. More broadly, they can include citizens whose opportunities and rights are impacted by institutional decisions [520]. As recipients of decisions, subjects typically have little control over the process unless they are also decision-makers. Yet, their feedback and experiences can shape decision-making indirectly, mediated through *developers*, *guardians*, or *decision-makers* who adjust AI systems or decision-making practices [504].

Knowledge providers (460). Knowledge providers supply data, expertise, or judgment that train and validate AI systems for decision-making. They may be domain experts (e.g., clinicians [26, 59, 98, 206, 272, 460, 539, 599] or legal professionals [284]), annotators, survey and experiment participants, or crowd workers [409]. Their contributions are often partial or imperfect [597], seeding model errors and uncertainty. Human cognition and biological mechanisms are also a source of inspiration for new model designs [41, 256, 327, 347], while human data serve as benchmarks for system performance [115, 197, 282, 608]. Knowledge providers can *influence* HAID through the AI systems trained, but often don't provide direct input (see Sec. 5.2 below).

Guardians (250). Guardians are individuals or institutions that establish rules, policies, and ethical guardrails for AI systems and decision-making processes. They include regulators [48, 291], policymakers [184, 291], and governments [437], ethicists and philosophers [142, 245], and research ethics committees [374]. Legal actors such as judges [69], courts [291], and parole boards [36], also act as guardians. *Input:* Guardians provide governance and oversight to distribute control over HAID, including setting guidelines [548] and monitoring systems [494].

Stakeholders (148). HAID also affects communities [288], citizens [561], company shareholders [490], vulnerable groups [226], and society at large [60, 546]. While these stakeholders typically do not provide direct input, their interests and concerns can *influence* the situational, cultural, and professional norms around *decision-makers*, subsequently incorporated into the decision-making process (see Sec. 5.2 below).

Developers (124). AI researchers and engineers design, build, and maintain AI systems [154]. They typically do not participate

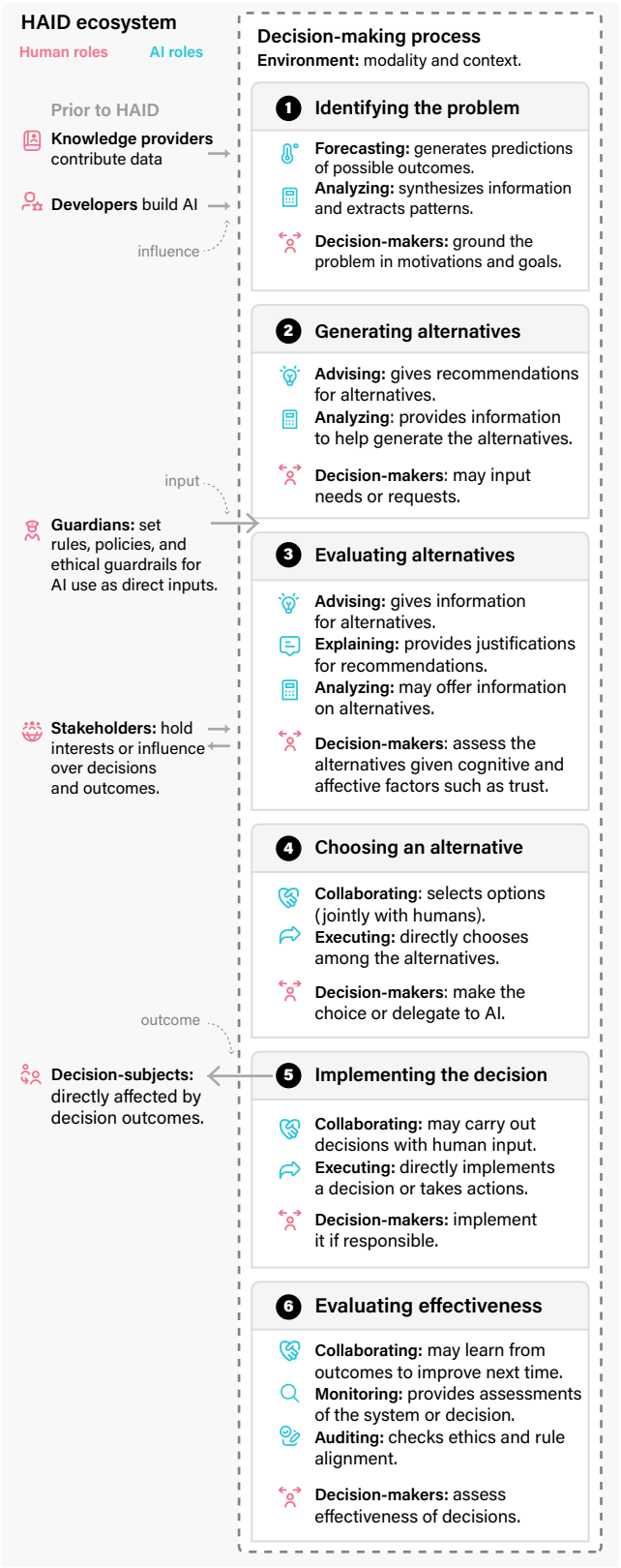


Figure 5: Position AI and human roles: *decision-makers* span all stages, AI roles often serve on specific stages, and a few human roles are outside the process.

directly in decision-making, but their design choices affect system reliability and performance, thereby *influencing* both the process and outcomes. Developers may collaborate with *decision-makers* [414, 614], *stakeholders* [81], *knowledge providers* [31, 410], and *guardians* [355, 408, 537, 596] to create and refine the systems, making them partially accountable.

4.3 Decision environment

The environments where humans and AI meet can shape how information is exchanged and presented, how humans perceive AI systems, and how human cognitive and affective traits come into play.

Modality. These range from textual [17, 314] and visual displays [267, 314, 440] to auditory [367] and 3D immersive formats [405, 535], and from conversational interfaces [91, 202] to interactive or semi-autonomous systems [196, 333]. In some cases, the environment is embodied in physical, wearable devices [227, 405, 609].

Decision context. Situational constraints, cultural factors, and social influences can shape decision-making. These factors may *concord* with existing information, such as group consensus [465], shared team information [204], peer-derived knowledge (e.g., clinical prescriptions from colleagues [395]), or peer recommendations [462]. They may also *discord*, as in conflicting peer review opinions [348], disruptive group behaviors [118], or structural challenges (e.g., workforce shortages [203]). Other overarching constraints further shape outcomes, including temporal pressures [404, 526], economic considerations [199, 404], and trade-offs between competing priorities (e.g., long-term outcomes vs. frequency [258]).

5 Consequences of HAID

While roles and their inputs define the architecture of HAID, the consequences of human–AI interactions in decision-making extend beyond immediate outcomes to shape human capabilities and system behaviors over time. To capture these dynamics, we examine the reciprocal *influence* that humans and AI exert on each other (see Fig. 6).

5.1 AI influence on humans

We observe 6 *AI influences* on decision-makers and the broader community. While some influences may be predicted from the HAID architecture, we focus on empirical effects observed in the surveyed papers.

Changing cognitive-affective states (N = 390). While cognitive-affective factors are preconditions to decision-making, they are also consequences of human–AI interactions.

Trust (291). remains central, closely tied to *reliance*. Trust operates as both a human input and a quality shaped by interaction with AI. It may be *behavioral* (reflected in actions like accepting AI advice [261, 571]), *cognitive* (based on reasoning [61]), or *affective* (linked to confidence or comfort [61, 349]). Well-designed explanations and interfaces can foster and calibrate trust [51, 55, 74, 195, 202, 321, 486, 501, 547, 571, 594], though they may lead to over-trust and misplaced reliance [51, 283, 314, 388]. Situational factors, such as framing and perceived accuracy, further complicate these dynamics [426]. Other manifestations include algorithm aversion [35, 180] or inflated faith in AI capabilities [261]. For a fuller account, see

dedicated surveys on trust [215, 553, 583].

Cognitive demands (201). This aspect concerns *mental effort* and *user engagement* in decision-making. Information from AI may increase cognitive load, especially when complex or overwhelming [4, 11, 151, 580, 617]. Example-based explanations may improve decision accuracy, but impose mental effort on users unfamiliar with technical details [151]. Conversely, AI can reduce cognitive load by translating complex information into digestible formats (e.g., *analyzing* role) [264, 275]. *Explaining* roles can improve user understanding of AI outputs [568], yet cause confusion or misinterpretation when outputs are incorrect, random, or unnecessarily complex [267, 493, 579]. Similarly, user engagement can be shaped by task and interface designs [322].

Affective or perceptual responses (187). AI can influence users' feelings, attitudes, emotions, and subjective perceptions in decision-making. It may boost *confidence* [366, 474, 477], as explanations can shift users' trust calibration and confidence [474, 359]. Yet confidence can drop [239, 531] when users are outvoted by AI teammates [239]. *Satisfaction* shows similar patterns. Some studies report higher satisfaction [367, 498, 517, 562], while others find reduced satisfaction when users feel misunderstood by AI [189]. Negative *emotions* also emerge: AI in cancer screening has been linked to heightened patient anxiety [403] and unnecessary information may harm well-being [32].

Altering decision outcomes (386). AI roles like *advising* and *explaining* target improving decision quality and outcomes, with many surveyed papers demonstrating success. AI involvement can enhance accuracy while reducing errors [74, 103, 159, 224, 302, 587], improve response speed and rates [62, 506], and achieve complementary human-AI performance [314, 465, 472]. A meta-analysis of human–AI collaboration [542] shows that AI often augments decision quality compared to humans alone, though human–AI teams may underperform AI operating independently. However, AI can diminish decision quality when providing incorrect advice [163, 249].

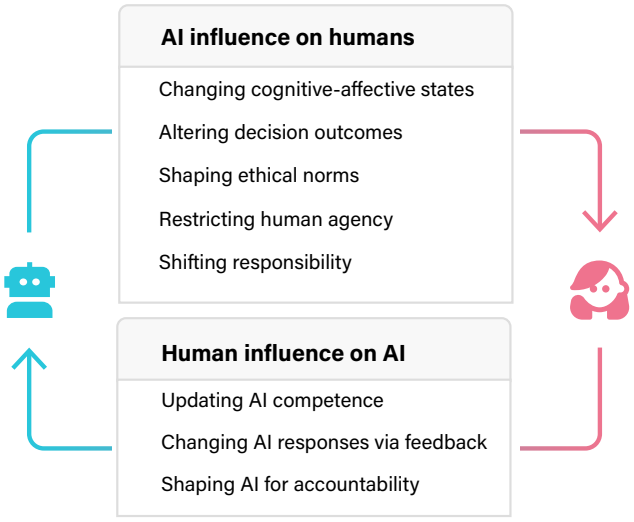


Figure 6: Summary of human and AI influences on each other: 5 AI influences and 3 human influences.

Shaping ethical norms (81). The use of AI in decision-making inevitably raises questions about equity, moral guidance, and privacy. Studies show AI can help reduce existing bias (e.g., gender) in hiring and other judgments [172, 205, 268, 436]. Yet AI systems also risk reproducing or amplifying existing biases [192, 365], or leading people to accept unethical advice when framed as objective or data-driven [589]. Such outcomes depend heavily on system architecture and design choices [90, 504]. Privacy concerns also loom large: systems that use sensitive data for profiling or prediction (e.g., credit scoring, healthcare) can erode personal control [429].

Restricting human agency (60). Surveyed papers reveal a tension between AI roles and human agency. Automation can reduce autonomy and dispossess agency [9, 284, 377, 431, 521]: in manufacturing training, full automation diminished workers' perceived autonomy, whereas partial automation maintained engagement and agency [431]. Conversely, interactive or adaptive designs can increase decision control [85], as group decision interfaces enable participants to negotiate and maintain influence over outcomes, improving their sense of agency [592].

Shifting responsibility (31). AI involvement can alter responsibility attribution for decisions and outcomes. People often redistribute or diffuse responsibility: joint decision-making with AI can reduce the extent to which humans are faulted [494]. Conversely, algorithmic systems may obscure accountability, leaving it unclear who should be held responsible when errors occur, or harm is caused [377, 589]. Yet design interventions, such as interface choices, can heighten perceived accountability [7].

5.2 Human influence on AI

Beyond serving as direct input in decision-making, human roles also influence AI systems. Compared to AI influence on humans, human influence on AI is less documented.

Updating AI competence (N = 141). Although *developers* and *knowledge providers* do not directly participate in HAID, they shape it through the competence of AI systems they build. *Developers'* design choices, priorities, and implementation strategies determine whether AI systems support or undermine decision-making [34]. Many surveyed papers propose design strategies to improve AI competence [412]. *Knowledge providers*—domain specialists, annotators, experiment participants, and crowd workers—contribute data and expertise underpinning AI systems. These contributions range from expert judgments and decisions [54, 284, 316, 450] to large-scale labels and ratings [627], complemented by contextual and qualitative resources such as documentation and case records [337].

Changing AI responses through feedback (108). Human users (often *decision-makers*) can directly shape or adjust AI outputs through clarifying, refining, correcting, or evaluating feedback. For example, reinforcement learning algorithms can learn from human feedback [502], clarification or confirmation can resolve uncertainty in generative AI responses [522], blind and low-vision users can query systems to refine AI-generated image descriptions [487], and binary positive or negative feedback can improve reinforcement learning efficiency [615]. Human experts may also directly correct AI-generated assignments or forecasts prior to decision-making [233, 271].

Shaping AI for accountability (56). *Stakeholders* and *guardians* can influence AI systems by enforcing norms, constraints, and oversight into their design, thereby enhancing accountability beyond guiding decision-making itself. This includes creating evaluation protocols [429] and responding to regulatory pressure to avoid protected characteristics in sensitive domains [550] and addressing legal considerations [265]. However, critical perspectives also warn against moral scapegoating, such as assigning responsibility to AI to shield human or institutional actors from accountability [494]. *Knowledge providers* may also contribute to moral, ethical, and social norms [277, 550].

6 Related Work: Surveys of HAID

Before turning to our discussion, we review related surveys to help situate our contributions, as several prior surveys have also addressed HAID directly.

The most relevant is by Lai et al. [300], who surveyed 7 conferences and collected 74 papers between 2018–2021. Their review focused on human-subjects experiments, categorizing decision domains, information provided, and evaluation metrics, while advocating for a decision space. We are inspired by their methodology but cover significantly broader ground, moving beyond human-subjects experiments (expanding, for instance, from 5 to 13 domains). Some issues they identified—such as limited exploration of design spaces and narrow decision tasks—have been at least partially addressed by our work, while others persist, notably the lack of benchmark data.

Two other highly relevant surveys evaluated trust in automated systems [553] and trust calibration [583]. While trust emerges as a central theme in our corpus, our review spans a broader landscape beyond trust considerations alone. Similarly, Salimzadeh et al. emphasized task complexity in HAID [466], and Vaccaro et al. conducted a meta-analysis on human–AI decision performance [542]. Within XAI-assisted decision-making, Bertrand et al. reviewed cognitive biases [38], and Schemmer et al. provided a comprehensive meta-analysis [473]. They represent valuable but more scoped efforts.

A much larger body of domain-focused surveys also exists. For instance, Hassan et al. surveyed patients' perceptions of AI decision-aids used for screening, prevention, prognosis, and treatment [214], while others reviewed machine learning for clinical decision support [53, 305, 433, 533, 549, 637]. Additional surveys touch on HAID within broader reviews [12, 40, 122, 146, 149, 278, 383, 451, 456]. These surveys illustrate both breadth and fragmentation. Our contribution lies in integrating general and domain-specific works to provide a broader and more coherent picture of HAID.

7 Discussion: Frontiers of HAID

We synthesize our findings to outline key research gaps and pathways for future work, organized around four perspectives: (1) building coherent foundations, (2) positioning the roles of humans and AI, (3) accounting for changes over time, and (4) breaking down research community silos. We close by reflecting on our systematic review process.

7.1 Building coherent foundations

One primary challenge for us is the comparability of the literature. The issues arise from both **a lack of uniform terminology** and **unclear boundaries** around what constitutes HAID—including fundamental questions about what human–AI decision-making means and where the line lies between HAID and what is not. For example, our *advising* role appears variously as "AI-assisted" [331, 332], "AI-aided"[20], "AI-augmented" [503], "AI-powered"[559], among others like "AI-mediated" [143] and "AI-induced" [620]. These terms signal some degree of AI involvement but rarely clarify assumptions about AI competence or expected human compliance, leaving critical questions unanswered: which decision stages involve AI? How do humans and AI interact? Even identical terms like "AI-assisted decision(-making)" carry different interpretations. While such inconsistency is expected for a new field drawing from different domains, it makes synthesizing knowledge about HAID particularly challenging. Our definition in Sec. 1.1 attempts to bring conceptual clarity to this field by articulating required parties and their contributions, clarifying *who* is involved and *how* their inputs enter the decision-making process. Our role taxonomy (Sec. 4) operationalizes this by classifying the specific functions AI and humans perform across decision stages. Together, these elements articulate (some) assumptions masked by terms like "AI-assisted" or "AI-augmented."

Similarly, we find minimal supporting Common Task Frameworks (CTF) [222]—community efforts to solve the same task with shared data and metrics—in the context of HAID, whether as benchmark datasets or standardized evaluation tasks. While criticized for encouraging incremental research and introducing shared biases [518], CTF enables comparative baselines and building cumulative knowledge. These same issues were noted in Lai et al.'s survey covering 2018–2021 [300]. Despite covering three additional years of research, we identified few advances (see Sec. 3.1 *dataset contributions*). New and human-centered benchmark datasets remain scarce. While defining standardized human tasks across domains is inherently difficult, we contend that HCI has successfully established common evaluation paradigms, such as those used in Fitts' Law [360] or visualization perception [101].

Suggestion 1. We call for the development of common decision tasks, reporting standards, and comparative benchmarks to systematically and reliably assess HAID research. Possible concrete tasks include establishing domain-specific benchmark datasets with standardized protocols (e.g., diagnostic reasoning steps) and defining minimal reporting metrics (e.g., confidence, reliance); our corpus and survey provide a rich source for both (Secs. 3.2 and 5). While task diversity across disciplines is real, shared *mechanisms* based on cognitive processes rather than domain expertise allow cross-domain benchmarks or tasks. Our use of 4 decision levels and the 6-step process model offers one organizing structure. Behavioral science provides additional foundations: frameworks for uncertainty [534], multi-stage [555], and sequential decisions [495] can be adapted as common task paradigms, much as Fitts' Law standardized tasks.

7.2 Positioning human and AI

Most works we surveyed emphasize only one or two human roles—for example, *decision-makers* in human-subjects studies (and in algorithmic works, even decision-makers were absent), or *guardians* in accountability discussions. Current evaluation methods may miss interactions between different human roles; little work has advocated for dynamics between *decision-subjects* and *decision-makers* [265]. **Suggestion 2. Meaningful progress in the field requires considering the ecosystem:** *decision-subjects*, *developers*, *stakeholders*, *guardians*, and *knowledge providers*. While it is not realistic for a single study to account for every role, researchers can broaden their perspective—for example, inviting human decision-makers to reflect on decision-subjects' perspectives, or coordinating with *developers* in user studies.

Suggestion 3. Future works could explicitly articulate the underlying assumptions about AI competence and examine these assumptions. Concretely, research could specify AI roles (e.g., *advising*) or the targeted stage(s) of decision-making (e.g., choosing an alternative), evaluate whether decision-makers' perceptions align, and report these design choices as metadata in publications. Such clarity will enable comparing results across studies, diagnosing why AI support succeeds or fails, and building cumulative knowledge. While our taxonomy offers an example of operationalization, alternative frameworks include considering the information gain in decision tasks to reduce task ambiguity [200, 201, 244]. In system or artifact development, researchers might further explicitly document whether a system supports the entire decision-making process or specific stages, and if the latter, how to connect with other stages.

7.3 Accounting for changes over time

We also feel dynamics in HAID warrant more attention. Similar concerns have been raised in recent surveys, such as critiques of focus on decision trials [300] or calls for adaptive strategies of trust calibration [583].

Suggestion 4. We should attend to the evolving nature of human and AI systems. Among surveyed papers, we find a striking scarcity: only about 15 reported adaptive AI (e.g., [24, 247, 408, 461]) and about 10 examined human adaptation (e.g., [49, 165, 408])—so few that these dimensions, originally planned for systematic extraction, yielded insufficient data for meaningful analysis. This is concerning because humans naturally reflect on their prior decisions and adapt strategies over time. For example, humans learn to write more effective prompts after seeing LLM outputs, while LLMs may adjust based on those prompts. This mutual adaptation makes it problematic to assume static cognitive or affective traits. Simultaneously, the development of adaptive AI systems is accelerating. Systems that adjust to user behavior, context, or goals become common, yet few user studies evaluate them realistically. While HCI researchers themselves need time to learn and experiment with newer systems, this lag means our understanding trails current technological capabilities. We therefore advocate for research investigating decision-making with dynamic or adaptive behaviors, such as longitudinal studies tracking decision quality and strategy shifts, and evaluation frameworks measuring co-adaptation. For example, quantifying the evolving competence gap between human and AI or measuring changes in AI literacy over time.

Suggestion 5. Human influence on AI systems deserves more attention. We found only three types of such influence (see Sec. 5.2). Even when AI appears to have *executing* roles, this autonomy typically stems from human delegation. Yet as AI research advances, the assumption of humans maintaining full control over system behavior is increasingly challenged. Recent discussions around LLM jailbreaks [576], safety training failures [576], and misalignment [39] highlight this tension. In the context of HAID, understanding the mechanisms and limits of human influence remains underexplored, representing an opportunity for HCI researchers to contribute, and example studies include empirical research documenting control failure modes and metrics of controllability across standardized tasks.

7.4 Breaking down research community silos

Despite shared interests, research on HAID remains fragmented across disciplinary boundaries, with HCI, AI, and domain-specific communities at times pursuing parallel questions in isolation. This fragmentation appears in three ways.

Decision tasks within the same domains are highly similar. The same databases contain highly similar papers addressing similar problems with similar framings: IEEE publications cluster around autonomous driving, while Elsevier and Springer Nature almost exclusively target medical Decision Support Systems or algorithmic diagnosis. ACM venues center on user studies of explanation and trust alongside fairness and ethical frameworks. While this concentration is understandable—conferences and journals define their scope, and researchers from the same domain share interests and build upon each other's work—the resulting work can appear overly similar, at times incremental, and in some cases indistinguishable. This insularity proves disciplinary silos that foster local depth but risk narrowing inquiry.

Cross-domain connections remain limited despite complementary strengths. AI/ML domains contribute theoretical foundations, algorithmic support, and generalizable frameworks [167, 251, 350]; medical and healthcare domains offer specialized datasets and access to professionals [312, 497, 544]; HCI provides rigorous user-centered methods and empirical knowledge; law and psychology contribute philosophical and normative discussions [284, 353, 629]. Despite promising connections [248, 621], these domains still largely operate in isolation. HCI researchers may lament the absence of generalizable frameworks that exist in AI/ML [300], while medical researchers may struggle with evaluation methods developed in HCI. This fragmentation is compounded by publication practices: domain-specific expectations make interdisciplinary research difficult to publish in a single domain.

Survey work often reinforces rather than bridges domain isolation. Researchers who review and synthesize a field remain within their familiar venues. Several previous surveys focus exclusively on ACM conferences [300, 553, 583], while others draw only from domain-specific databases (e.g., PubMed [242], Medline [42]). While cross-domain surveys require substantial effort and risk rejection from disciplinary venues that may find such breadth unfocused, we feel that this practice perpetuates the very fragmentation it should help address.

Suggestion 6. Individual researchers and groups can seek collaborators outside their disciplinary communities and include literature from other domains. While structural solutions at the level of publishing or funding require broad institutional coordination, smaller initiatives can offer immediate possibilities: cross-conference workshops, interdisciplinary seminars, and shared reading groups could build bridges across fields incrementally. These grassroots efforts may ultimately lead to institutional changes.

7.5 The role of this work

Our work takes a step toward addressing the challenges outlined above. We connect previously fragmented fields of HAID by surveying across 20 databases and publishers, unifying them under shared taxonomies. While we do not claim comprehensive coverage of this rapidly evolving field, we demonstrate the feasibility of cross-field synthesis and showcase a method for conducting a cross-domain survey while remaining thematically focused. We hope this serves multiple audiences who are interested in this field: providing **beginner researchers** a broad map of possibilities, encouraging **junior researchers** to explore emerging frontiers, and inspiring **senior researchers** to reflect on the field.

Our corpus and synthesis also open new directions. First, researchers can surface new research questions by viewing how different fields approach HAID, discovering untested assumptions or overlooked variables. Second, we identify many important factors (e.g., agency, affective factors) that merit deeper research; while meta-analyses of each were beyond our scope, our corpus provides a resource for relevant papers. Finally, our work maps out decision tasks within a particular domain and across domains, creating opportunities for new theories of HAID and policy development.

7.6 Reflection

Looking back, we acknowledge that our work is not without limitations. Given our bandwidth, we did not include arXiv papers or the most recent 2025 publications, though new work continues to emerge [201, 244]. We adopted one decision-making process model, but recognize that people sometimes rely on heuristics [178], bounded rationality [257], or domain-specific decision frameworks [124] that our model may not fully capture. Future work may benefit from comparing findings across multiple theoretical lenses, such as dual-process theory [148] and naturalistic decision-making [645].

We also reflect on methodological lessons learned. We combined LLM-based venue selection, SVM classification, LLM-assisted coding, and custom annotation interfaces. In hindsight, this pipeline reveals both affordances and constraints. LLM-based venue selection reduces uncertainty as our authors lack comprehensive venue knowledge to reliably classify venues across disciplines, and the included HCI and AI/ML venues generally align with our expectations. The two-stage screening (Sec. 2.4) was adopted to address limitations from the second iteration, but proved less efficient than anticipated. A single-stage manual approach used in our validation would demand comparable effort with fewer missed papers. Comparing LLM outputs with our conventionally-coded subset, long contexts (full-text analysis, long prompts) introduce drift, and LLMs generate labels beyond the codebook. We informed subsequent coders about these biases and required two rounds of verification, with final codes determined through consensus among human coders. This approach helps bound bias within acceptable margins for large-scale analysis. While LLM-assisted coding shows promise, with only one human coder, further exploration remains a task for future work.

Finally, we want to pause and ask: what constitutes HCI research in an era of ubiquitous AI? As we stepped outside familiar venues published within ACM to examine HAID across computer science, medicine, law, and psychology, we encountered a wealth of literature and perspectives far beyond our initial expectations. This breadth enabled us to assemble a more comprehensive picture of human–AI decision-making—our landscape. Was this effort worthwhile? From our perspective, yes. We were at times overwhelmed, but also rewarded and genuinely excited by the knowledge we gained. This experience reminds us not to approach research with presumptions about what we might find, but rather with openness.

Author Contributions

Yixin Bai contributed to data collection, developed the taxonomies, conducted the majority of the qualitative coding, and contributed to manuscript editing. Taehyun Yang and Zhongzheng Xu served as major secondary coders throughout the study, each coding approximately half of the corpus, and contributed to manuscript editing; Taehyun Yang additionally served as the independent coder. Dayeon Ki, Ziyi Wang, and Yu Hou served as secondary coders for the initial submission and coded a subset of the final corpus. Fumeng Yang contributed to data collection, developed parts of the taxonomies, managed the qualitative analysis, drafted the manuscript, and conceived and supervised the research.

References

[1] 2025. CORE Conference Portal. https://www.core.edu.au/conference-portal.

[2] 2025. Resurchify: Impact Score Search Engine. https://www.resurchify.com/if/impact-factor-search.

[3] 2025. Web of Science Journal Info (WoS-Journal.info). https://wos-journal.info/.

[4] Ammar N. Abbas, Chidera W. Amazu, Joseph Mietkiewicz, Houda Briwa, Andres Alonso Perez, Gabriele Baldissone, Micaela Demichela, Georgios C. Chasparis, John D. Kelleher, and Maria Chiara Leva. 2024. Analyzing operator states and the impact of AI-enhanced decision support in control rooms: a human-in-the-loop specialized reinforcement learning framework for intervention strategies. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2391605 [CORPUS].

[5] Babak Abbasi, Toktam Babaei, Zahra Hosseinifard, Kate Smith-Miles, and Maryam Dehghani. 2020. Predicting solutions of large-scale optimization problems via machine learning: a case study in blood supply chain management. *Computers & Operations Research* (2020). doi:10.1016/j.cor.2020.104941 [CORPUS].

[6] Joanna Abraham, Brian Bartek, Alicia Meng, Christopher Ryan King, Bing Xue, Chenyang Lu, and Michael S. Avidan. 2023. Integrating machine learning predictions for perioperative risk management: towards an empirical design of a flexible-standardized risk assessment tool. *Journal of Biomedical Informatics* (2023). doi:10.1016/j.jbi.2022.104270 [CORPUS].

[7] Martin Adam. 2022. Accountability-based user interface design artifacts and their implications for user acceptance of AI-enabled services. In *European Conference on Information Systems*. [CORPUS].

[8] Deepesh Agarwal and Balasubramaniam Natarajan. 2023. Tracking and handling behavioral biases in active learning frameworks. *Information Sciences* (2023). doi:10.1016/j.ins.2023.119117 [CORPUS].

[9] Mariyani Ahmad Husairi and Patricia Rossi. 2024. Delegation of purchasing tasks to AI: The role of perceived choice and decision autonomy. *Decision Support Systems* (2024). doi:10.1016/j.dss.2023.114166 [CORPUS].

[10] Hadi Akbarian, Farhad Mahmoudi Jalali, Mohammad Gheibi, Mostafa Hajiaghaei-Keshteli, Mehran Akrami, and Ajit K. Sarmah. 2022. A sustainable decision support system for soil bioremediation of toluene incorporating UN sustainable development goals. *Environmental Pollution* (2022). doi:10.1016/j.envpol.2022.119587 [CORPUS].

[11] Kamala Aliyeva and Nijat Mehdiyev. 2024. Uncertainty-aware multi-criteria decision analysis for evaluation of explainable artificial intelligence methods: a use case from the healthcare domain. *Information Sciences* (2024). doi:10.1016/j.ins.2023.119987 [CORPUS].

[12] Azhar Alsufyani, Omer Rana, and Charith Perera. 2024. Knowledge-based Cyber Physical Security at Smart Home: A Review. *Comput. Surveys* (2024). doi:10.1145/3698768 .

[13] Rohan Alur, Loren Laine, Darrick Li, Manish Raghavan, Devavrat Shah, and Dennis Shung. 2023. Auditing for human expertise. In *Advances in Neural Information Processing Systems*. [CORPUS].

[14] American Psychological Association. 2018. *Decision-making*.

[15] Asbjørn Ammitzbøll Flügge, Thomas Hildebrandt, and Naja Holten Møller. 2021. Street-Level Algorithms and AI in Bureaucratic Decision-Making: A Caseworker Perspective. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3449114 [CORPUS].

[16] Haozhe An, Christabel Acquaye, Colin Wang, Zongxia Li, and Rachel Rudinger. 2024. Do Large Language Models Discriminate in Hiring Decisions on the Basis of Race, Ethnicity, and Gender?. In *Annual Meeting of the Association for Computational Linguistics (Short Papers)*. doi:10.18653/v1/2024.acl-short.37 [CORPUS].

[17] Naomi Aoki. 2021. The importance of the assurance that ""humans are still in the decision loop"" for public trust in artificial intelligence: evidence from an online experiment. *Computers in Human Behavior* (2021). doi:10.1016/j.chb.2020.106572 [CORPUS].

[18] S. Arena, E. Florian, I. Zennaro, P. F. Orrù, and F. Sgarbossa. 2022. A novel decision support system for managing predictive maintenance strategies based on machine learning approaches. *Safety Science* (2022). doi:10.1016/j.ssci.2021.105529 [CORPUS].

[19] Ines Arous, Jie Yang, Mourad Khayati, and Philippe Cudre-Mauroux. 2021. Peer Grading the Peer Reviews: A Dual-Role Approach for Lightening the Scholarly Paper Review Process. In *Proceedings of the International World Wide Web Conference (The Web Conference)*. doi:10.1145/3442381.3450088 [CORPUS].

[20] David Askay, Anuraj Dhillon, and Lynn Metcalf. 2024. Achieving decisional fit with AI-aided group decisions: The role of intuitive decision-making style in predicting perceived fairness and decision acceptance. In *Hawaii International Conference on System Sciences*. [CORPUS].

[21] Abdullah M. Baabdullah. 2024. The precursors of AI adoption in business: Towards an efficient decision-making and functional performance. *International Journal of Information Management* (2024). doi:10.1016/j.ijinfomgt.2023.102745 [CORPUS].

[22] Jayachandu Bandlamudi, Kushal Mukherjee, Prerna Agarwal, Sampath Dechu, Siyu Huo, Vatche Isahagian, Vinod Muthusamy, Naveen Purushothaman, and Renuka Sindhgatta. 2024. Towards Hybrid Automation by Bootstrapping Conversational Interfaces for IT Operation Tasks. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 13 (Jul. 2024), 15654–15660. doi:10.1609/aaai.v37i13.26856 [CORPUS].

[23] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S. Weld, Walter S. Lasecki, and Eric Horvitz. 2019. Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Tradeoff. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v33i01.33012429 [CORPUS].

[24] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the Whole Exceed its Parts? The Effect of AI Explanations on Complementary Team Performance. Article 81, 16 pages. doi:10.1145/3411764.3445717 [CORPUS].

[25] Chelsea Barabas, Madars Virza, Karthik Dinakar, Joichi Ito, and Jonathan Zittrain. 2018. Interventions over predictions: reframing the ethical debate for actuarial risk assessment. In *Conference on Fairness, Accountability, and Transparency (FAccT)*. [CORPUS].

[26] Catarina Barata, Veronica Rotemberg, Noel C. F. Codella, Philipp Tschandl, Christoph Rinner, Bengu Nisa Akay, Zoe Apalla, Giuseppe Argenziano, Allan Halpern, Aimilios Lallas, Caterina Longo, Josep Malvehy, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. 2023. A reinforcement learning model for AI-based decision support in skin cancer. *Nature Medicine* (2023). doi:10.1038/s41591-023-02475-5 [CORPUS].

[27] Alina Jade Barnett, Fides Regina Schwartz, Chaofan Tao, Chaofan Chen, Yinhao Ren, Joseph Y. Lo, and Cynthia Rudin. 2021. A case-based interpretable deep learning model for classification of mass lesions in digital mammography. *Nature Machine Intelligence* (2021). doi:10.1038/s42256-021-00423-x [CORPUS].

[28] George Baryannis, Samir Dani, and Grigoris Antoniou. 2019. Predicting supply chain risks using machine learning: the trade-off between performance and interpretability. *Future Generation Computer Systems* (2019). doi:10.1016/j.future.2019.07.059 [CORPUS].

[29] Edmon Begoli, Tanmoy Bhattacharya, and Dimitri Kusnezov. 2019. The need for uncertainty quantification in machine-assisted medical decision making. *Nature Machine Intelligence* 1 (2019), 20–23. doi:10.1038/s42256-018-0004-1 .

[30] Vahid Behzadan, James Minton, and Arslan Munir. 2019. TrolleyMod v1.0: An Open-Source Simulation and Data-Collection Platform for Ethical Decision Making in Autonomous Vehicles. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3306618.3314239 [CORPUS].

[31] Amine Belhadi, Sachin Kamble, Samuel Fosso Wamba, and Maciel M. Queiroz. 2022. Building supply-chain resilience: an artificial intelligence-based technique and decision-making framework. *International Journal of Production Research* (2022). doi:10.1080/00207543.2021.1950935 [CORPUS].

[32] Dennis Benner, Sofia Marlena Sch¨obel, Andreas Janson, and Jan Marco Leimeister. 2022. How to achieve ethical persuasive design: a review and theoretical propositions for information systems. *AIS Transactions on Human-Computer Interaction* (2022). [CORPUS].

[33] Nina Corvelo Benz and Manuel Rodriguez. 2023. Human-aligned calibration for AI-assisted decision making. In *Advances in Neural Information Processing Systems*. [CORPUS].

[34] Arngeir Berge, Frode Guribye, Siri-Linn Schmidt Fotland, Gro Fonnes, Ingrid H. Johansen, and Christoph Trattner. 2023. Designing for Control in Nurse-AI Collaboration During Emergency Medical Calls. In *Proceedings of the ACM Designing Interactive Systems Conference*. doi:10.1145/3563657.3596110 [CORPUS].

[35] Benedikt Berger, Martin Adam, Alexander R¨uhr, and Alexander Benlian. 2021. Watch me improve—algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering* (2021). [CORPUS].

[36] Richard Berk. 2017. An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology* (2017). doi:10.1007/s11292-017-9286-2 [CORPUS].

[37] Michael H. Bernstein, Michael K. Atalay, Elizabeth H. Dibble, Aaron W. P. Maxwell, Adib R. Karam, Saurabh Agarwal, Robert C. Ward, Terrance T. Healey, and Grayson L. Baird. 2023. Can incorrect artificial intelligence (AI) results impact radiologists, and if so, what can we do about it? A multi-reader pilot study of lung cancer detection with chest radiography. *European Radiology* (2023). doi:10.1007/s00330-023-09747-1 [CORPUS].

[38] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534164 .

[39] Jan Betley, Daniel Tan, Niels Warncke, Anna Sztyber-Betley, Xuchan Bao, Martín Soto, Nathan Labenz, and Owain Evans. 2025. Emergent Misalignment: Narrow finetuning can produce broadly misaligned LLMs. arXiv:2502.17424 [cs.CL] https://arxiv.org/abs/2502.17424

[40] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3461702.3462571 .

[41] Raunak Bhattacharyya, Blake Wulfe, Derek J. Phillips, Alex Kuefler, Jeremy Morton, Ransalu Senanayake, and Mykel J. Kochenderfer. 2023. Modeling Human Driving Behavior Through Generative Adversarial Imitation Learning. *IEEE Transactions on Intelligent Transportation Systems* (2023). doi:10.1109/TITS.2022.3227738 [CORPUS].

[42] Amanda Bianco, Zaid A.M. Al-Azzawi, Elena Guadagno, Esli Osmanlliu, Jocelyn Gravel, and Dan Poenaru. 2023. Use of machine learning in pediatric surgical clinical prediction tools: a systematic review. *Journal of Pediatric Surgery* (2023). doi:10.1016/j.jpedsurg.2023.01.020 .

[43] Ioana Bica, Daniel Jarrett, Alihan Hüyük, and Mihaela van der Schaar. 2021. Learning ""what-if"" explanations for sequential decision-making. In *International Conference on Learning Representations*. [CORPUS].

[44] Nadine Bienefeld, Jens Michael Boss, Rahel L¨uthy, Dominique Brodbeck, Jan Azzati, Mirco Blaser, Jan Willms, and Emanuela Keller. 2023. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. *Nature Partner Journals Digital Medicine* (2023). doi:10.1038/s41746-023-00837-4 [CORPUS].

[45] Ivo Blohm, Torben Antretter, Charlotta Sirén, Dietmar Grichnik, and Joakim Wincent. 2022. It's a peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice* (2022). doi:10.1177/1042258720945206 [CORPUS].

[46] Marko Bohanec, Mirjana Kljajić Borštnar, and Marko Robnik-Šikonja. 2017. Explaining machine learning models in sales predictions. *Expert Systems with Applications* (2017). doi:10.1016/j.eswa.2016.11.010 [CORPUS].

[47] Douglas Bosse, Steven Thompson, and Peter Ekman. 2023. In consilium apparatus: artificial intelligence, stakeholder reciprocity, and firm performance. *Journal of Business Research* (2023). doi:10.1016/j.jbusres.2022.113402 [CORPUS].

[48] Kiel Brennan-Marquez and Stephen E. Henderson. 2019. Artificial intelligence and role-reversible judgment. *The Journal of Criminal Law and Criminology* (2019). doi:10.2307/48572776 [CORPUS].

[49] Noelle Brown, Koriann South, and Eliane S. Wiese. 2022. The Shortest Path to Ethics in AI: An Integrated Assignment Where Human Concerns Guide Technical Decisions. In *ACM Conference on International Computing Education Research (ICER)*. doi:10.1145/3501385.3543978 [CORPUS].

[50] Magda Bucholc, Xuemei Ding, Haiying Wang, David H. Glass, Hui Wang, Girijesh Prasad, Liam P. Maguire, Anthony J. Bjourson, Paula L. McClean, Stephen Todd, David P. Finn, and KongFatt Wong-Lin. 2019. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert Systems with Applications* (2019). doi:10.1016/j.eswa.2019.04.022 [CORPUS].

[51] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3449287 [CORPUS].

[52] Adrian Bumann. 2023. No ground truth at sea – developing high-accuracy AI decision-support for complex environments. In *Hawaii International Conference on System Sciences*. [CORPUS].

[53] Michał Burdukiewicz, Jarosław Chilimoniuk, Krystyna Grzesiak, Adam Kręstowski, and Michał Ciborowski. 2024. Ml-based clinical decision support models based on metabolomics data. *TrAC Trends in Analytical Chemistry* (2024). doi:10.1016/j.trac.2024.117819 .

[54] Eleanor R. Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuścińska, Suzanne Currie, J. Marc Overhage, Erika S. Poole, and Jofish Kaye. 2023. Healthcare AI Treatment Decision Support: Design Principles to Enhance Clinician Adoption and Trust. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581251 [CORPUS].

[55] Zana Buçinca, Phoebe Lin, Krzysztof Z. Gajos, and Elena L. Glassman. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3377325.3377498 [CORPUS].

[56] Yewon Byun, Dylan Sam, Michael Oberst, Zachary Lipton, and Bryan Wilder. 2024. Auditing fairness under unobserved confounding. In *The International Conference on Artificial Intelligence and Statistics (AISTATS)*. [CORPUS].

[57] Ángel Alexander Cabrera, Adam Perer, and Jason I. Hong. 2023. Improving Human-AI Collaboration With Descriptions of AI Behavior. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3579612 [CORPUS].

[58] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Conference on Human Factors in Computing Systems*. doi:10.1145/3290605.3300234 [CORPUS].

[59] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. Hello AI: Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* (2019). doi:10.1145/3359206 [CORPUS].

[60] Han Cai, Kan Ren, Weinan Zhang, Kleanthis Malialis, Jun Wang, Yong Yu, and Defeng Guo. 2017. Real-Time Bidding by Reinforcement Learning in Display Advertising. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*. doi:10.1145/3018661.3018702 [CORPUS].

[61] Jingyuan Cai and Fiona Nah. 2024. User Acceptance of Advice by AI Agents: Expectation-System Fit Perspective. In *International Conference on Information Systems*. [CORPUS].

[62] Agostina Calabrese, Leonardo Neves, Neil Shah, Maarten Bos, Björn Ross, Mirella Lapata, and Francesco Barbieri. 2024. Explainability and Hate Speech: Structured Explanations Make Social Media Moderators Faster. In *Annual Meeting of the Association for Computational Linguistics (Short Papers)*. doi:10.18653/v1/2024.acl-short.38 [CORPUS].

[63] Francisco Maria Calisto, Nuno Nunes, and Jacinto C. Nascimento. 2022. Modeling adoption of intelligent agents in medical imaging. *International Journal of Human-Computer Studies* (2022). doi:10.1016/j.ijhcs.2022.102922 [CORPUS].

[64] Gabriele Campanella, Matthew G. Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck Krauss Silva, Klaus J. Busam, Edi Brogi, Victor E. Reuter, David S. Klimstra, and Thomas J. Fuchs. 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature Medicine* (2019). doi:10.1038/s41591-019-0508-1 [CORPUS].

[65] Cindy Candrian and Anne Scherer. 2022. Rise of the machines: Delegating decisions to autonomous AI. *Computers in Human Behavior* (2022). doi:10.1016/j.chb.2022.107308 [CORPUS].

[66] Burak Cankaya, Kazim Topuz, Dursun Delen, and Aaron Glassman. 2023. Evidence-based managerial decision-making with machine learning: the case of bayesian inference in aviation incidents. *Omega - The International Journal of Management Science* (2023). doi:10.1016/j.omega.2023.102906 [CORPUS].

[67] Shiye Cao, Anqi Liu, and Chien-Ming Huang. 2024. Designing for Appropriate Reliance: The Roles of AI Uncertainty Presentation, Initial User Decision, and User Demographics in AI-Assisted Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* (2024). doi:10.1145/3637318 [CORPUS].

[68] Kathleen M Carley and Dean Behrens. 1999. Organizational and individual decision making. *Handbook of systems engineering and management* 1 (1999).

[69] Zachariah Carmichael and Walter J. Scheirer. 2023. Unfooling Perturbation-Based Post Hoc Explainers. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v37i6.25847 [CORPUS].

[70] G. Castañé, A. Dolgui, N. Kousi, B. Meyers, S. Thevenin, E. Vyhmeister, and P-O. Östberg. 2023. The assistant project: AI for high level decisions in manufacturing. *International Journal of Production Research* 0, 0 (2023), 1–20. doi:10.1080/00207543.2022.2069525 [CORPUS].

[71] Federico Maria Cau, Hanna Hauptmann, Lucio Davide Spano, and Nava Tintarev. 2023. Effects of AI and Logic-Style Explanations on Users' Decisions Under Different Levels of Uncertainty. *ACM Transactions on Interactive Intelligent Systems* (2023). doi:10.1145/3588320 [CORPUS].

[72] Lenart Celar and Ruth M. J. Byrne. 2023. How people reason with counterfactual and causal explanations for Artificial Intelligence decisions in familiar and unfamiliar domains. *Memory & Cognition* (2023). doi:10.3758/s13421-023-01407-5 [CORPUS].

[73] Joymallya Chakraborty, Suvodeep Majumder, and Tim Menzies. 2021. Bias in Machine Learning Software: Why? How? What to Do?. In *Proceedings of the ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE)*. doi:10.1145/3468264.3468537 [CORPUS].

[74] Tirtha Chanda, Katja Hauser, Sarah Hobelsberger, Tabea-Clara Bucher, Carina Nogueira Garcia, Christoph Wies, Harald Kittler, Philipp Tschandl, Cristian Navarrete-Dechent, Sebastian Podlipnik, Emmanouil Chousakos, Iva Crnaric, Jovana Majstorovic, Linda Alhajwan, Tanya Foreman, Sandra Peternel, Sergei Sarap, İrem Ü̈zdemir, Raymond L. Barnhill, Mar Llamas-Velasco, Gabriela Poch, Sören Korsing, Wiebke Sondermann, Frank Friedrich Gellrich, Markus V. Heppt, Michael Erdmann, Sebastian Haferkamp, Konstantin Drexler, Matthias Goebeler, Bastian Schilling, Jochen S. Utikal, Kamran Ghoreschi, Stefan Fröhling, Eva Krieghoff-Henning, Titus J. Brinker, and Reader Study Consortium. 2024. Dermatologist-like explainable AI enhances trust and confidence in diagnosing melanoma. *Nature Communications* (2024). doi:10.1038/s41467-023-43095-4 [CORPUS].

[75] Eshwar Chandrasekharan, Chaitrali Gandhi, Matthew Wortley Mustelier, and Eric Gilbert. 2019. Crossmod: A Cross-Community Learning-based System to Assist Reddit Moderators. *Proceedings of the ACM on Human-Computer Interaction* (2019). doi:10.1145/3359276 [CORPUS].

[76] Shruthi Chari, Oshani Seneviratne, Daniel M. Gruen, Morgan A. Foreman, Amar K. Das, and Deborah L. McGuinness. 2020. Explanation Ontology: A Model of Explanations for User-Centered AI. In *International Semantic Web Conference*. doi:10.1007/978-3-030-62466-8_15 [CORPUS].

[77] Veena Chattaraman, Wi-Suk Kwon, Kassandra Ross, Jihyun Sung, Kiana Alikhademi, Brianna Richardson, and Juan E. Gilbert. 2024. 'Smart' choice? Evaluating ai-based mobile decision bots for in-store decision-making. *Journal of Business Research* (2024). doi:10.1016/j.jbusres.2024.114801 [CORPUS].

[78] Sheshadri Chatterjee, Ranjan Chaudhuri, and Demetris Vrontis. 2022. AI and digitalization in relationship management: impact of adopting AI-embedded CRM system. *Journal of Business Research* (2022). doi:10.1016/j.jbusres.2022.06.033 [CORPUS].

[79] Haoyu Chen, Wenbin Lu, Rui Song, and Pulak Ghosh. 2024. On learning and testing of counterfactual fairness through data preprocessing. *J. Amer. Statist. Assoc.* (2024). doi:10.1080/01621459.2023.2186885 [CORPUS].

[80] Hannah MD Chen, Xiaoyue MS Ma, Hal BS Rives, Aisha Serpedin, Peter MD Yao, and MPhil MS FACS Rameau, Anaïs MD. 2024. Trust in machine learning driven clinical decision support tools among otolaryngologists. *The Laryngoscope* (2024). doi:10.1002/lary.31260 [CORPUS].

[81] Sikai Chen, Jiqian Dong, Paul (Young Joun) Ha, Yujie Li, and Samuel Labi. 2021. Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles. *Computer-Aided Civil and Infrastructure Engineering* (2021). doi:10.1111/mice.12702 [CORPUS].

[82] Valerie Chen, Q. Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. 2023. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3610219 [CORPUS].

[83] Xia Chen and Philipp Geyer. 2022. Machine assistance in energy-efficient building design: a predictive framework toward dynamic interaction with human decision-making under uncertainty. *Applied Energy* (2022). doi:10.1016/j.apenergy.2021.118240 [CORPUS].

[84] Xiusi Chen, Jyun-Yu Jiang, Kun Jin, Yichao Zhou, Mingyan Liu, P. Jeffrey Brantingham, and Wei Wang. 2022. ReLiable: Offline Reinforcement Learning for Tactical Strategies in Professional Basketball Games. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. doi:10.1145/3511808.3557105 [CORPUS].

[85] Yifang Chen, Alex Cuellar, Haipeng Luo, Jignesh Modi, Heramb Nemlekar, and Stefanos Nikolaidis. 2020. Fair contextual multi-armed bandits: theory and experiments. In *Conference on Uncertainty in Artificial Intelligence (UAI)*. [CORPUS].

[86] Yufei Chen, Xiaodong Yue, Hamido Fujita, and Siyuan Fu. 2017. Three-way decision support for diagnosis on focal liver lesions. *Knowledge-Based Systems* (2017). doi:10.1016/j.knosys.2017.04.008 [CORPUS].

[87] Zhiyu Chen, Jason Ingyu Choi, Besnik Fetahu, and Shervin Malmasi. 2024. Identifying High Consideration E-Commerce Search Queries. In *Empirical Methods in Natural Language Processing (EMNLP)*. doi:10.18653/v1/2024.emnlp-industry.42 [CORPUS].

[88] Furui Cheng, Dongyu Liu, Fan Du, Yanna Lin, Alexandra Zytek, Haomin Li, Huamin Qu, and Kalyan Veeramachaneni. 2022. VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models. *IEEE Transactions on Visualization and Computer Graphics* (2022). doi:10.1109/TVCG.2021.3114836 [CORPUS].

[89] Furui Cheng, Yao Ming, and Huamin Qu. 2021. DECE: Decision Explorer with Counterfactual Explanations for Machine Learning Models. *IEEE Transactions on Visualization and Computer Graphics* (2021). doi:10.1109/TVCG.2020.3030342 [CORPUS].

[90] Hao-Fei Cheng, Logan Stapleton, Anna Kawakami, Venkatesh Sivaraman, Yanghuidi Cheng, Diana Qing, Adam Perer, Kenneth Holstein, Zhiwei Steven Wu, and Haiyi Zhu. 2022. How Child Welfare Workers Reduce Racial Disparities in Algorithmic Decisions. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3501831 [CORPUS].

[91] Chun-Wei Chiang, Zhuoran Lu, Zhuoyan Li, and Ming Yin. 2024. Enhancing AI-Assisted Group Decision Making through LLM-Powered Devil's Advocate. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3640543.3645199 [CORPUS].

[92] Hana Chockler and Joseph Y. Halpern. 2022. On Testing for Discrimination Using Causal Models. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v36i5.20494 [CORPUS].

[93] Leah Chong, Guanglu Zhang, Kosa Goucher-Lambert, Kenneth Kotovsky, and Jonathan Cagan. 2022. Human confidence in artificial intelligence and in themselves: the evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior* (2022). doi:10.1016/j.chb.2021.107018 [CORPUS].

[94] Amit K. Chopra and Munindar P. Singh. 2018. Sociotechnical Systems and Ethics in the Large. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3278721.3278740 [CORPUS].

[95] Alexandra Chouldechova, Diana Benavides-Prado, Oleksandr Fialko, and Rhema Vaithianathan. 2018. A case study of algorithm-assisted decision making in child maltreatment hotline screening decisions. In *Conference on Fairness, Accountability, and Transparency (FAccT)*. [CORPUS].

[96] Hyesun Choung, John S. Seberger, and Prabu David. 2024. When AI is Perceived to be Fairer than a Human: Understanding Perceptions of Algorithmic Decisions in a Job Application Context. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2023.2266244 [CORPUS].

[97] Diego Chowell, Seong-Keun Yoo, Cristina Valero, Alessandro Pastore, Chirag Krishna, Mark Lee, Douglas Hoen, Hongyu Shi, Daniel W. Kelly, Neal Patel, Vladimir Makarov, Xiaoxiao Ma, Lynda Vuong, Erich Y. Sabio, Kate Weiss, Fengshen Kuo, Tobias L. Lenz, Robert M. Samstein, Nadeem Riaz, Prasad S. Adusumilli, Vinod P. Balachandran, George Plitas, A. Ari Hakimi, Omar Abdel-Wahab, Alexander N. Shoushtari, Michael A. Postow, Robert J. Motzer, Marc Ladanyi, Ahmet Zehir, Michael F. Berger, Mithat Gönen, Luc G. T. Morris, Nils Weinhold, and Timothy A. Chan. 2022. Improved prediction of immune checkpoint blockade efficacy across multiple cancer types. *Nature Biotechnology* (2022). doi:10.1038/s41587-021-01070-8 [CORPUS].

[98] Patrick Ferdinand Christ, Mohamed Ezzeldin A. Elshaer, Florian Ettlinger, Sunil Tatavarty, Marc Bickel, Patrick Bilic, Markus Rempfler, Marco Armbruster, Felix Hofmann, Melvin D'Anastasi, Wieland H. Sommer, Seyed-Ahmad Ahmadi, and Bjoern H. Menze. 2016. Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. doi:10.1007/978-3-319-46723-8_48 [CORPUS].

[99] Alton Y.K. Chua, Anjan Pal, and Snehasish Banerjee. 2023. AI-enabled Investment Advice: Will Users Buy It? *Computers in Human Behavior* (2023). doi:10.1016/j.chb.2022.107481 [CORPUS].

[100] Ilker Cingillioglu. 2024. What impacts matriculation decisions? Identifying students' university choice factors on a global scale with artificial intelligence. *Studies in Higher Education* (2024). doi:10.1080/03075079.2024.2319870 [CORPUS].

[101] William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *J. Amer. Statist. Assoc.* 79, 387 (1984), 531–554. doi:10.1080/01621459.1984.10478080

[102] Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh. 2021. Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3442188.3445921 [CORPUS].

[103] Julien Colin, Thomas FEL, Remi Cadene, and Thomas Serre. 2022. What I Cannot Predict, I Do Not Understand: A Human-Centered Evaluation Framework for Explainability Methods. In *Advances in Neural Information Processing Systems*. [CORPUS].

[104] Joe Collenette, Katie Atkinson, and Trevor Bench-Capon. 2023. Explainable AI Tools for Legal Reasoning about Cases: A Study on the European Court of Human Rights. *Artificial Intelligence* (2023). doi:10.1016/j.artint.2023.103861 https://doi.org/10.1016/j.artint.2023.103861.

[105] Sam Corbett-Davies, Johann D. Gaebler, Hamed Nilforoshan, Ravi Shroff, and Sharad Goel. 2023. The measure and mismeasure of fairness. *Journal of Machine Learning Research* (2023). [CORPUS].

[106] Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini. 2022. Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3531146.3533213 [CORPUS].

[107] Mutlu Cukurova, Carmel Kent, and Rosemary Luckin. 2019. Artificial intelligence and multimodal data in the service of human decision-making: a case study in debate tutoring. *British Journal of Educational Technology* (2019). doi:10.1111/bjet.12829 [CORPUS].

[108] Geoff Currie, Adrien-Maxence Hespel, and Ann Carstens. 2023. Australian Perspectives on Artificial Intelligence in Veterinary Practice. *Veterinary Radiology & Ultrasound* (2023). doi:10.1111/vru.13234 [CORPUS].

[109] Xinyue Dai, Mark T. Keane, Laurence Shalloo, Elodie Ruelle, and Ruth M.J. Byrne. 2022. Counterfactual Explanations for Prediction and Diagnosis in XAI. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534144 [CORPUS].

[110] Xiang Dai, Maciej Rybinski, and Sarvnaz Karimi. 2021. SearchEHR: A Family History Search System for Clinical Decision Support. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. doi:10.1145/3459637.3481986 [CORPUS].

[111] Alexander D'Amour, Hansa Srinivasan, James Atwood, Pallavi Baljekar, D. Sculley, and Yoni Halpern. 2020. Fairness is not static: deeper understanding of long term fairness via simulation studies. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3351095.3372878 [CORPUS].

[112] Anupam Datta, Shayak Sen, and Yair Zick. 2016. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. In *Proceedings of the IEEE Symposium on Security and Privacy*. doi:10.1109/SP.2016.42 [CORPUS].

[113] Hans de Bruijn, Martijn Warnier, and Marijn Janssen. 2022. The perils and pitfalls of explainable AI: strategies for explaining algorithmic decision-making. *Government Information Quarterly* (2022). doi:10.1016/j.giq.2021.101666 .

[114] Francisco M. De La Vega, Shimul Chowdhury, Barry Moore, Erwin Frise, Jeanette McCarthy, Edgar Javier Hernandez, Terence Wong, Kiely James, Lucia Guidugli, Pankaj B. Agrawal, Casie A. Genetti, Catherine A. Brownstein, Alan H. Beggs, Britt-Sabina L¨oscher, Andre Franke, Braden Boone, Shawn E. Levy, Katrin ¨Ounap, Sander Pajusalu, Matt Huentelman, Keri Ramsey, Marcus Naymik, Vinodh Narayanan, Narayanan Veeraraghavan, Paul Billings, Martin G. Reese, Mark Yandell, and Stephen F. Kingsmore. 2021. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. *Genome Medicine* (2021). doi:10.1186/s13073-021-00965-0 [CORPUS].

[115] Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, Francesco Orciuoli, and Enrique Herrera-Viedma. 2016. A framework for context-aware heterogeneous group decision making in business processes. *Knowledge-Based Systems* (2016). doi:10.1016/j.knosys.2016.03.019 [CORPUS].

[116] Oscar Blessed Deho, Chen Zhan, Jiuyong Li, Jixue Liu, Lin Liu, and Thuc Duy Le. 2022. How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics? *British Journal of Educational Technology* (2022). doi:10.1111/bjet.13217 [CORPUS].

[117] Yashar Deldjoo, Vito Walter Anelli, Hamed Zamani, Alejandro Bellogín, and Tommaso Noia. 2021. A flexible framework for evaluating user and item fairness in recommender systems. *User Modeling and User-Adapted Interaction* (2021). doi:10.1007/s11257-020-09285-1 [CORPUS].

[118] Amra Delić, Hanif Emamgholizadeh, Francesco Ricci, and Judith Masthoff. 2024. Supporting Group Decision-Making: Insights from a Focus Group Study. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. doi:10.1145/3627043.3659538 [CORPUS].

[119] Quirin Demlehner and Sven Laumer. 2020. Shall we use it or not? Explaining the adoption of artificial intelligence for car manufacturing purposes. In *European Conference on Information Systems*. [CORPUS].

[120] Quirin Demlehner and Sven Laumer. 2024. How the Terminator Might Affect the Car Manufacturing Industry: Examining the Role of Pre-Announcement Bias for AI-Based IS Adoptions. *Information & Management* (2024). doi:10.1016/j.im.2023.103881 [CORPUS].

[121] Qilin Deng, Hao Li, Kai Wang, Zhipeng Hu, Runze Wu, Linxia Gong, Jianrong Tao, Changjie Fan, and Peng Cui. 2021. Globally Optimized Matchmaking in Online Games. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3447548.3467074 [CORPUS].

[122] Brian T. Denton. 2023. Frontiers of medical decision-making in the modern age of data analytics. *Institute of Industrial and Systems Engineers Transactions* (2023). doi:10.1080/24725854.2022.2092918 .

[123] Michael Desmond, Michael Muller, Zahra Ashktorab, Casey Dugan, Evelyn Duesterwald, Kristina Brimijoin, Catherine Finegan-Dollak, Michelle Brachman, Aabhas Sharma, Narendra Nath Joshi, and Qian Pan. 2021. Increasing the Speed

and Accuracy of Data Labeling Through an AI Assisted Interface. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3397481.3450698 [CORPUS].

[124] Dennis J. Devine and Steve W.J. Kozlowski. 1995. Domain-Specific Knowledge and Task Characteristics in Decision Making. *Organizational Behavior and Human Decision Processes* 64, 3 (1995), 294–306. doi:10.1006/obhd.1995.1107

[125] Christian Dietzmann and Yanqing Duan. 2022. Artificial Intelligence for Managerial Information Processing and Decision-Making in the Era of Information Overload. In *Hawaii International Conference on System Sciences*. [CORPUS].

[126] Murat Dikmen and Catherine Burns. 2022. The effects of domain knowledge on trust in explainable AI and task performance: a case of peer-to-peer lending. *International Journal of Human-Computer Studies* (2022). doi:10.1016/j.ijhcs.2022.102792 [CORPUS].

[127] Boris Djartov, Sanaz Mostaghim, Anne Papenfuß, and Matthias Wies. 2024. A Learning Classifier System Approach to Time-Critical Decision-Making in Dynamic Alternate Airport Selection. In *IEEE Congress on Evolutionary Computation*. doi:10.1109/CEC60901.2024.10612016 [CORPUS].

[128] Abhishek Djeachandrane, Said Hoceini, Serge Delmas, Jean-Michel Duquerrois, Alain Dubois, and Abdelhamid Mellouk. 2022. Deep RL-based Abnormal Behavior Detection and Prevention in Network Video Surveillance. In *IEEE Global Communications Conference*. doi:10.1109/GLOBECOM48099.2022.10001471 [CORPUS].

[129] Abhishek Djeachandrane, Said Hoceini, Serge Delmas, Jean-Michel Duquerrois, and Abdelhamid Mellouk. 2022. QoE-based Situational Awareness-Centric Decision Support for Network Video Surveillance. In *IEEE International Conference on Communications*. doi:10.1109/ICC45855.2022.9838601 [CORPUS].

[130] Vishwas Dohale, Milind Akarte, Angappa Gunasekaran, and Priyanka Verma. 2024. Exploring the role of artificial intelligence in building production resilience: Learnings from the COVID-19 pandemic. *International Journal of Production Research* (2024). doi:10.1080/00207543.2022.2127961 [CORPUS].

[131] Hamidreza Ahady Dolatsara, Eyyub Kibis, Musa Caglar, Serhat Simsek, Ali Dag, Gelareh Ahadi Dolatsara, and Dursun Delen. 2022. An interpretable decision-support systems for daily cryptocurrency trading. *Expert Systems with Applications* (2022). doi:10.1016/j.eswa.2022.117409 [CORPUS].

[132] Kate Donahue, Alexandra Chouldechova, and Krishnaram Kenthapadi. 2022. Human-Algorithm Collaboration: Achieving Complementarity and Avoiding Unfairness. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3531146.3533221 [CORPUS].

[133] Kate Donahue, Sreenivas Gollapudi, and Kostas Kollias. 2024. When Are Two Lists Better than One?: Benefits and Harms in Joint Decision-Making. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v38i9.28866 [CORPUS].

[134] Daniel A. D¨oppner, Patrick Derckx, and Detlef Schoder. 2018. An intelligent decision support system for the empty unit load device repositioning problem in air cargo industry. In *Hawaii International Conference on System Sciences*. [CORPUS].

[135] Julia Dressel and Hany Farid. 2018. The accuracy, fairness, and limits of predicting recidivism. *Science Advances* (2018). [CORPUS].

[136] Iddo Drori and Dov Te'eni. 2024. Human-in-the-loop AI reviewing: feasibility, opportunities, and risks. *Journal of the Association for Information Systems* (2024). [CORPUS].

[137] Krishnamurthy (Dj) Dvijotham, Jim Winkens, Melih Barsbey, Sumedh Ghaisas, Robert Stanforth, Nick Pawlowski, Patricia Strachan, Zahra Ahmed, Shekoofeh Azizi, Yoram Bachrach, Laura Culp, Mayank Daswani, Jan Freyberg, Christopher Kelly, Atilla Kiraly, Timo Kohlberger, Scott McKinney, Basil Mustafa, Vivek Natarajan, Krzysztof Geras, Jan Witowski, Zhi Zhen Qin, Jacob Creswell, Shravya Shetty, Marcin Sieniek, Terry Spitz, Greg Corrado, Pushmeet Kohli, Taylan Cemgil, and Alan Karthikesalingam. 2023. Enhancing the reliability and accuracy of AI-enabled diagnosis via complementarity-driven deferral to clinicians. *Nature Medicine* 29, 7 (2023), 1814–1820. doi:10.1038/s41591-023-02437-x [CORPUS].

[138] Sjur Dyrkolbotn, Truls Pedersen, and Marija Slavkovik. 2018. On the Distinction between Implicit and Explicit Ethical Agency. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3278721.3278769 [CORPUS].

[139] Karin Eberhard. 2023. The effects of visualization on judgment and decision-making: a systematic literature review. *Management Review Quarterly* 73, 1 (2023), 167–214.

[140] Jessica Maria Echterhoff, Aditya Melkote, Sujen Kancherla, and Julian McAuley. 2024. Avoiding Decision Fatigue with AI-Assisted Decision-Making. In *ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. doi:10.1145/3627043.3659569 [CORPUS].

[141] Jessica Maria Echterhoff, Matin Yarmand, and Julian McAuley. 2022. AI-Moderated Decision-Making: Capturing and Balancing Anchoring Bias in Sequential Decision Tasks. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3517443 [CORPUS].

[142] Linda Eggert. 2023. Autonomised harming. *Philosophical Studies* (2023). doi:10.1007/s11098-023-01990-y [CORPUS].

[143] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445188 [CORPUS].

[144] Jon Eklof, Ulrika Snis, Thomas Hamelryck, Alexander Grima, and Ola Ronning. 2024. AI Implementation and Capability Development in Manufacturing: An Action Research Case. In *Hawaii International Conference on System Sciences*. [CORPUS].

[145] Hannah Elder, Casey Canfield, Daniel B. Shank, Tobias Rieger, and Casey Hines. 2024. Knowing when to pass: the effect of AI reliability in risky decision contexts. *Human Factors: The Journal of the Human Factors and Ergonomics Society* (2024). doi:10.1177/00187208221100691 [CORPUS].

[146] Sandra Eloranta and Magnus Boman. 2022. Predictive models for clinical decision making: deep dives in practical machine learning. *Journal of Internal Medicine* (2022). doi:10.1111/joim.13483 .

[147] Alexander Erlei, Richeek Das, Lukas Meub, Avishek Anand, and Ujwal Gadiraju. 2022. For What It's Worth: Humans Overwrite Their Economic Self-interest to Avoid Bargaining With AI Systems. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3517734 [CORPUS].

[148] Jonathan St. B. T. Evans and Keith E. Stanovich. 2013. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science* 8, 3 (2013), 223–241. doi:10.1177/1745691612460685 PMID: 26172965.

[149] Katherine J. Evans, Andrew Terhorst, and Byeong Ho Kang. 2017. From data to decisions: helping crop producers build their actionable knowledge. *Critical Reviews in Plant Sciences* (2017). doi:10.1080/07352689.2017.1336047 .

[150] Cristiano Fabbri, Michele Lombardi, Enrico Malaguti, and Michele Monaci. 2024. On-line strategy selection for reducing overcrowding in an emergency department. *Omega - The International Journal of Management Science* (2024). doi:10.1016/j.omega.2024.103098 [CORPUS].

[151] Tobias Benjamin Fahse, Ivo Blohm, and Benjamin van Giffen. 2022. Effectiveness of example-based explanations to improve human decision quality in machine learning forecasting systems. In *International Conference on Information Systems*. [CORPUS].

[152] Xudong Fan, Xijin Zhang, and Xiong (Bill) Yu. 2022. A graph convolution network-deep reinforcement learning model for resilient water distribution network repair decisions. *Computer-Aided Civil and Infrastructure Engineering* (2022). doi:10.1111/mice.12813 [CORPUS].

[153] Carlos Fernández-Loría, Foster Provost, and Xintian Han. 2022. Explaining data-driven decisions made by AI systems: the counterfactual approach. *Management Information Systems Quarterly* (2022). [CORPUS].

[154] Carmine Ferrara, Giulia Sellitto, Filomena Ferrucci, Fabio Palomba, and Andrea De Lucia. 2023. Fairness-aware machine learning engineering: how far are we? *Empirical Software Engineering* (2023). doi:10.1007/s10664-023-10402-y [CORPUS].

[155] S. Foersch, M. Eckstein, D.-C. Wagner, F. Gach, A.-C. Woerl, J. Geiger, C. Glasner, S. Schelbert, S. Schulz, S. Porubsky, A. Kreft, A. Hartmann, A. Agaimy, and W. Roth. 2021. Deep Learning for Diagnosis and Survival Prediction in Soft Tissue Sarcoma. *Annals of Oncology* (2021). doi:10.1016/j.annonc.2021.06.007 [CORPUS].

[156] Riccardo Fogliato, Shreya Chappidi, Paul Lungren, Matthew; Fisher, Diane Wilson, Michael Fitzke, Mark Parkinson, Kori Horvitz, Eric; Inkpen, and Besmira Nushi. 2022. Who Goes First? Influences of Human-AI Workflow on Decision Making in Clinical Imaging. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3531146.3533193 [CORPUS].

[157] Raymond Fok, Nedim Lipka, Tong Sun, and Alexa F Siu. 2024. Marco: Supporting Business Document Workflows via Collection-Centric Information Foraging with Large Language Models. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3641969 [CORPUS].

[158] Ângelo Fonseca, Axel Ferreira, Luís Ribeiro, Sandra Moreira, and Cristina Duque. 2024. Embracing the future—is artificial intelligence already better? A comparative study of artificial intelligence performance in diagnostic accuracy and decision-making. *European Journal of Neurology* (2024). doi:10.1111/ene.16195 [CORPUS].

[159] Laura Fontanesi, Sebastian Gluth, Mikhail S. Spektor, and J¨org Rieskamp. 2019. A reinforcement learning diffusion decision model for value-based decisions. *Psychonomic Bulletin & Review* (2019). doi:10.3758/s13423-018-1554-2 [CORPUS].

[160] Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John Dickerson, and Vincent Conitzer. 2018. Adapting a Kidney Exchange Algorithm to Align With Human Values. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v32i1.11505 [CORPUS].

[161] Elena Freisinger and Sabrina Schneider. 2021. Only a coward hides behind AI? Preferences in surrogate, moral decision-making. In *International Conference on Information Systems*. [CORPUS].

[162] Yuchuan Fu, Changle Li, Fei Richard Yu, Tom H. Luan, and Yao Zhang. 2020. A Decision-Making Strategy for Vehicle Autonomous Braking in Emergency via Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology* (2020). doi:10.1109/TVT.2020.2986005 [CORPUS].

[163] Andreas Fuegener, J¨orn Grahl, Alok Gupta, and Wolfgang Ketter. 2021. Will humans-in-the-loop become borgs? Merits and pitfalls of working with AI. *Management Information Systems Quarterly* (2021). [CORPUS].

[164] Saadia Gabriel, Liang Lyu, James Siderius, Marzyeh Ghassemi, Jacob Andreas, and Asuman E. Ozdaglar. 2024. MisinfoEval: Generative AI in the Era of "Alternative Facts". In *Empirical Methods in Natural Language Processing*. doi:10.18653/v1/2024.emnlp-main.487 [CORPUS].

[165] Krzysztof Z. Gajos and Lena Mamykina. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3490099.3511138 [CORPUS].

[166] Jie Gao, Yuchen Guo, Gionnieve Lim, Tianqin Zhang, Zheng Zhang, Toby Jia-Jun Li, and Simon Tangi Perrault. 2024. CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642002 [CORPUS].

[167] Ruijiang Gao, Maytal Saar-Tsechansky, Maria De-Arteaga, Ligong Han, Min Kyung Lee, and Matthew Lease. 2021. Human-ai collaboration with bandit feedback. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2021/237 [CORPUS].

[168] Xin Gao, Tian Luan, Xueyuan Li, Qi Liu, Zhaoyang Ma, Xiaoqiang Meng, and Zirui Li. 2024. Ethical Alignment Decision-Making for Connected Autonomous Vehicle in Traffic Dilemmas via Reinforcement Learning From Human Feedback. *IEEE Internet of Things Journal* (2024). doi:10.1109/JIOT.2024.3447070 [CORPUS].

[169] Xin Gao, Tian Luan, Xueyuan Li, Qi Liu, Xiaoqiang Meng, and Zirui Li. 2023. A Human Feedback-Driven Decision-Making Method Based on Multi-Modal Deep Reinforcement Learning in Ethical Dilemma Traffic Scenarios. In *IEEE International Conference on Intelligent Transportation Systems*. doi:10.1109/ITSC57777.2023.10422393 [CORPUS].

[170] Vael Gates, Thomas L. Griffiths, and Anca D. Dragan. 2020. How to be helpful to multiple people at once. In *Proceedings of the Cognitive Science Society*. doi:10.1111/cogs.12841 [CORPUS].

[171] Susanne Gaube, Harini Suresh, Martina Raue, Alexander Merritt, Seth J. Berkowitz, Eva Lermer, Joseph F. Coughlin, John V. Guttag, Errol Colak, and Marzyeh Ghassemi. 2021. Do as AI say: susceptibility in deployment of clinical decision-aids. *Nature Partner Journals Digital Medicine* (2021). doi:10.1038/s41746-021-00385-9 [CORPUS].

[172] Michael Geden and Joshua Andrews. 2021. Fair and Interpretable Algorithmic Hiring using Evolutionary Many Objective Optimization. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v35i17.17737 [CORPUS].

[173] Marcel Gehrung, Mireia Crispin-Ortuzar, Adam G. Berman, Maria O'Donovan, Rebecca C. Fitzgerald, and Florian Markowetz. 2021. Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning. *Nature Medicine* (2021). doi:10.1038/s41591-021-01287-9 [CORPUS].

[174] Meric Altug Gemalmaz and Ming Yin. 2022. Understanding Decision Subjects' Fairness Perceptions and Retention in Repeated Interactions with AI-Based Decision Systems. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534201 [CORPUS].

[175] Tomer Geva and Maytal Saar-Tsechansky. 2016. Who's a good decision maker? Data-driven expert worker ranking under unobservable quality. In *International Conference on Information Systems*.

[176] Tomer Geva and Maytal Saar-Tsechansky. 2020. Who is a better decision maker? Data-driven expert ranking under unobserved quality. *Production and Operations Management Society (POMS) Journal* 29, 8 (2020), 1971–1992. doi:10.1111/poms.13260 [CORPUS].

[177] Sarah N. Giest and Bram Klievink. 2024. More than a Digital System: How AI is Changing the Role of Bureaucrats in Different Organizational Contexts. *Public Management Review* (2024). doi:10.1080/14719037.2022.2095001 [CORPUS].

[178] Gerd Gigerenzer and Wolfgang Gaissmaier. 2011. Heuristic Decision Making. *Annual Review of Psychology* 62, Volume 62, 2011 (2011), 451–482. doi:10.1146/annurev-psych-120709-145346

[179] Zohar Gilad, Ofra Amir, and Liat Levontin. 2021. The Effects of Warmth and Competence Perceptions on Users' Choice of an AI System. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3446863 [CORPUS].

[180] Andrej Gill, Robert M. Gillenkirch, Julia Ortner, and Louis Velthuis. 2024. Dynamics of reliance on algorithmic advice. *Journal of Behavioral Decision Making* (2024). doi:10.1002/bdm.2414 [CORPUS].

[181] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning. In *2018 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. 80–89. doi:10.1109/DSAA.2018.00018

[182] Xavier Gitiaux and Huzefa Rangwala. 2019. Mdfa: multi-differential fairness auditor for black box classifiers. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2019/814 [CORPUS].

[183] Felipe Giuste, Wenqi Shi, Yuanda Zhu, Tarun Naren, Monica Isgut, Ying Sha, Li Tong, Mitali Gupte, and May D. Wang. 2023. Explainable Artificial Intelligence Methods in Combating Pandemics: A Systematic Review. *IEEE Reviews in Biomedical Engineering* (2023). doi:10.1109/RBME.2022.3185953 .

[184] David Glynn, John Giardina, Julia Hatamyar, Ankur Pandya, Marta Soares, and Noemi Kreif. 2024. Integrating decision modeling and machine learning to inform treatment stratification. *Health Economics* (2024). doi:10.1002/hec.4834 [CORPUS].

[185] Elisabeth Victoria Goessinger, Johannes-Christian Niederfeilner, Sara Cerminara, Julia-Tatjana Maul, Lisa Kostner, Michael Kunz, Stephanie Huber, Emrah Koral, Lea Habermacher, Gianna Sabato, Andrea Tadic, Carmina Zimmermann, Alexander Navarini, and Lara Valeska Maul. 2024. Patient and dermatologists' perspectives on augmented intelligence for melanoma screening: a prospective study. *Journal of the European Academy of Dermatology and Venereology* (2024). doi:10.1111/jdv.19905 [CORPUS].

[186] Grace Y. Gombolay, Andrew Silva, Mariah Schrum, Nakul Gopalan, Jamika Hallman-Cooper, Monideep Dutt, and Matthew Gombolay. 2024. Effects of explainable artificial intelligence in neurology decision support. *Annals of Clinical and Translational Neurology* (2024). doi:10.1002/acn3.52036 [CORPUS].

[187] Oscar Gomez, Steffen Holter, Jun Yuan, and Enrico Bertini. 2021. AdViCE: Aggregated Visual Counterfactual Explanations for Machine Learning Model Validation. In *IEEE Visualization Conference*. doi:10.1109/VIS49827.2021.9623271 [CORPUS].

[188] H.J. Gómez-Vallejo, B. Uriel-Latorre, M. Sande-Meijide, B. Villamarín-Bello, R. Pavón, F. Fdez-Riverola, and D. Glez-Peña. 2016. A case-based reasoning system for aiding detection and classification of nosocomial infections. *Decision Support Systems* (2016). doi:10.1016/j.dss.2016.02.005 [CORPUS].

[189] Ana Rita Gonçalves, Diego Costa Pinto, Saleh Shuqair, Marlon Dalmoro, and Anna S. Mattila. 2024. Artificial intelligence vs. autonomous decision-making in streaming platforms: a mixed-method approach. *International Journal of Information Management* (2024). doi:10.1016/j.ijinfomgt.2023.102748 [CORPUS].

[190] Bryce Goodman. 2021. Hard Choices and Hard Limits in Artificial Intelligence. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3461702.3462539 [CORPUS].

[191] Rajesh Govindan and Tareq Al-Ansari. 2019. Computational decision framework for enhancing resilience of the energy, water and food nexus in risky environments. *Renewable and Sustainable Energy Reviews* (2019). doi:10.1016/j.rser.2019.06.015 [CORPUS].

[192] Navita Goyal, Connor Baumler, Tin Nguyen, and Hal Daumé III. 2024. The Impact of Explanations on Fairness in Human-AI Decision-Making: Protected vs Proxy Features. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3640543.3645210 [CORPUS].

[193] Ben Green and Yiling Chen. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction* (2019). doi:10.1145/3359152 [CORPUS].

[194] Nina Grgić-Hlača, Muhammad Bilal Zafar, Krishna P. Gummadi, and Adrian Weller. 2018. Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v32i1.11296 [CORPUS].

[195] Stephan Grimmelikhuijsen. 2022. Explaining why the computer says no: algorithmic transparency affects the perceived trustworthiness of automated decision-making. *Public Administration Review* (2022). doi:10.1111/puar.13483 [CORPUS].

[196] Matthew Groh, Omar Badri, Roxana Daneshjou, Arash Koochek, Caleb Harris, Luis R. Soenksen, P. Murali Doraiswamy, and Rosalind Picard. 2024. Deep learning-aided decision support for diagnosis of skin disease across skin tones. *Nature Medicine* (2024). doi:10.1038/s41591-023-02728-3 [CORPUS].

[197] Mengzhuo Guo, Qingpeng Zhang, Xiuwu Liao, Frank Youhua Chen, and Daniel Dajun Zeng. 2021. A Hybrid Machine Learning Framework for Analyzing Human Decision-Making through Learning Preferences. *Omega - The International Journal of Management Science* (2021). doi:10.1016/j.omega.2020.102263 [CORPUS].

[198] Shunan Guo, Fan Du, Sana Malik, Eunyee Koh, Sungchul Kim, Zhicheng Liu, Donghyun Kim, Hongyuan Zha, and Nan Cao. 2019. Visualizing Uncertainty and Alternatives in Event Sequence Predictions. In *Conference on Human Factors in Computing Systems*. doi:10.1145/3290605.3300803 [CORPUS].

[199] Xin Guo, Muhammad Arslan Khalid, Ivo Domingos, Anna Lito Michala, Moses Adriko, Candia Rowel, Diana Ajambo, Alice Garrett, Shantimoy Kar, Xiaoxiang Yan, Julien Reboud, Edridah M. Tukahebwa, and Jonathan M. Cooper. 2021. Smartphone-based DNA diagnostics for malaria detection using deep learning for local decision support and blockchain technology for security. *Nature Electronics* (2021). doi:10.1038/s41928-021-00612-x [CORPUS].

[200] Ziyang Guo, Yifan Wu, Jason Hartline, and Jessica Hullman. 2024. Information Gain in Human-AI Collaboration. In *NeurIPS 2024 Workshop on Behavioral Machine Learning*. https://openreview.net/forum?id=p7a08mENAo

[201] Ziyang Guo, Yifan Wu, Jason Hartline, and Jessica Hullman. 2025. The Value of Information in Human-AI Decision-making. arXiv:2502.06152 [cs.AI] https://arxiv.org/abs/2502.06152

[202] Akshit Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. 2022. To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System. In *ACM Web Conference (formerly The World Wide Web Conference)*. doi:10.1145/3485447.3512248 [CORPUS].

[203] Shivam Gupta, Sachin Modgil, Régis Meissonier, and Yogesh K. Dwivedi. 2024. Artificial Intelligence and Information System Resilience to Cope With Supply Chain Disruption. *IEEE Transactions on Engineering Management* (2024). doi:10.1109/TEM.2021.3116770 [CORPUS].

[204] Necdet Gurkan and Bei Yan. 2023. Chatbot catalysts: improving team decision-making through cognitive diversity and information elaboration. In *International Conference on Information Systems*. [CORPUS].

[205] Felix Haag, Carlo Stingl, Katrin Zerfass, Konstantin Hopf, and Thorsten Staake. 2023. Overcoming anchoring bias: The potential of AI and XAI-based decision support. In *International Conference on Information Systems*. [CORPUS].

[206] H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. Ben Hadj Hassen, L. Thomas, A. Enk, L. Uhlmann, Christina Alt, Monika Arenbergerova, Renato Bakos, Anne Baltzer, Ines Bertlich, Andreas Blum, Therezia Bokor-Billmann, Jonathan Bowling, Naira Braghiroli, Ralph Braun, Kristina Buder-Bakhaya, Timo Buhl, Horacio Cabo, Leo Cabrijan, Naciye Cevic, Anna Classen, David Deltgen, Christine Fink, Ivelina Georgieva, Lara-Elena Hakim-Meibodi, Susanne Hanner, Franziska Hartmann, Julia Hartmann, Georg Haus, Elti Hoxha, Raimonds Karls, Hiroshi Koga, Jürgen Kreusch, Aimilios Lallas, Pawel Majenka, Ash Marghoob, Cesare Massone, Lali Mekokishvili, Dominik Mestel, Volker Meyer, Anna Neuberger, Kari Nielsen, Margaret Oliviero, Riccardo Pampena, John Paoli, Erika Pawlik, Barbar Rao, Adriana Rendon, Teresa Russo, Ahmed Sadek, Kinga Samhaber, Roland Schneiderbauer, Anissa Schweizer, Ferdinand Toberer, Lukas Trennheuser, Lyobomira Vlahova, Alexander Wald, Julia Winkler, Priscila Wölbing, and Iris Zalaudek. 2018. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology* (2018). doi:10.1093/annonc/mdy166 [CORPUS].

[207] H.A. Haenssle, C. Fink, F. Toberer, J. Winkler, W. Stolz, T. Deinlein, R. Hofmann-Wellenhof, A. Lallas, S. Emmert, T. Buhl, M. Zutt, A. Blum, M.S. Abassi, L. Thomas, I. Tromme, P. Tschandl, A. Enk, A. Rosenberger, Christina Alt, Marie Bachelerie, Sonali Bajaj, Alise Balcere, Sophie Baricault, Clément Barthaux, Yvonne Beckenbauer, Ines Bertlich, Andreas Blum, Marie-France Bouthenet, Sophie Brassat, Philipp Marcel Buck, Kristina Buder-Bakhaya, Maria-Letizia Cappelletti, Cécile Chabbert, Julie De Labarthe, Eveline DeCoster, Teresa Deinlein, Michèle Dobler, Daphnée Dumon, Steffen Emmert, Julie Gachon-Buffet, Mikhail Gusarov, Franziska Hartmann, Julia Hartmann, Anke Herrmann, Isabelle Hoorens, Eva Hulstaert, Raimonds Karls, Andreea Kolonte, Christian Kromer, Aimilios Lallas, Céline Le Blanc Vasseux, Annabelle Levy-Roy, Pawel Majenka, Marine Marc, Veronique Martin Bourret, Nadège Michelet-Brunacci, Christina Mitteldorf, Jean Paroissien, Camille Picard, Diana Plise, Valérie Reymann, Fabrice Ribeaudeau, Pauline Richez, Hélène Roche Plaine, Deborah Salik, Elke Sattler, Sarah Schäfer, Roland Schneiderbauer, Thierry Secchi, Karen Talour, Lukas Trennheuser, Alexander Wald, Priscila Wölbing, and Pascale Zukervar. 2020. Man against machine reloaded: performance of a market-approved convolutional neural network in classifying a broad spectrum of skin lesions in comparison with 96 dermatologists working under less artificial conditions. *Annals of Oncology* (2020). doi:10.1016/j.annonc.2019.10.013 [CORPUS].

[208] Holger Andreas Haenssle, Julia Katharina Winkler, Christine Fink, Ferdinand Toberer, Alexander Enk, Wilhelm Stolz, Teresa Deinlein, Rainer Hofmann-Wellenhof, Harald Kittler, Philipp Tschandl, Cliff Rosendahl, Aimilios Lallas, Andreas Blum, Mohamed Souhayel Abassi, Luc Thomas, Isabelle Tromme, Albert Rosenberger, Marie Bachelerie, Sonali Bajaj, Alise Balcere, Sophie Baricault, Clément Barthaux, Yvonne Beckenbauer, Ines Bertlich, Andreas Blum, Marie-France Bouthenet, Sophie Brassat, Philipp Marcel Buck, Kristina Buder-Bakhaya, Maria-Letizia Cappelletti, Cécile Chabbert, Julie De Labarthe, Eveline DeCoster, Teresa Deinlein, Michèle Dobler, Daphnée Dumon, Steffen Emmert, Julie Gachon-Buffet, Mikhail Gusarov, Franziska Hartmann, Julia Hartmann, Anke Herrmann, Isabelle Hoorens, Eva Hulstaert, Raimonds Karls, Andreea Kolonte, Christian Kromer, Aimilios Lallas, Céline Le Blanc Vasseux, Annabelle Levy-Roy, Pawel Majenka, Marine Marc, Veronique Martin Bourret, Nadège Michelet-Brunacci, Christina Mitteldorf, Jean Paroissien, Camille Picard, Diana Plise, Valérie Reymann, Fabrice Ribeaudeau, Pauline Richez, Hélène Roche Plaine, Deborah Salik, Elke Sattler, Sarah Schäfer, Roland Schneiderbauer, Thierry Secchi, Karen Talour, Lukas Trennheuser, Alexander Wald, Priscila Wölbing, and Pascale Zukervar. 2021. Skin lesions of face and scalp – classification by a market-approved convolutional neural network in comparison with 64 dermatologists. *European Journal of Cancer* (2021). doi:10.1016/j.ejca.2020.11.034 [CORPUS].

[209] Paul Hager, Friederike Jungmann, Robbie Holland, Kunal Bhagat, Inga Hubrecht, Manuel Knauer, Jakob Vielhauer, Marcus Makowski, Rickmer Braren, Georgios Kaissis, and Daniel Rueckert. 2024. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nature Medicine* (2024). doi:10.1038/s41591-024-03097-1 [CORPUS].

[210] Marina Haliem, Ganapathy Mani, Vaneet Aggarwal, and Bharat Bhargava. 2021. A Distributed Model-Free Ride-Sharing Approach for Joint Matching, Pricing, and Dispatching Using Deep Reinforcement Learning. *IEEE Transactions on Intelligent Transportation Systems* (2021). doi:10.1109/TITS.2021.3096537 [CORPUS].

[211] Ronan Hamon, Henrik Junklewitz, Gianclaudio Malgieri, Paul De Hert, Laurent Beslay, and Ignacio Sanchez. 2021. Impossible Explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3442188.3445917 [CORPUS].

[212] Benjamin V. Hanrahan, Anita Chen, JiaHua Ma, Ning F. Ma, Anna Squicciarini, and Saiph Savage. 2021. The Expertise Involved in Deciding which HITs are Worth Doing on Amazon Mechanical Turk. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3449202 [CORPUS].

[213] Qianyue Hao, Wenzhen Huang, Fengli Xu, Kun Tang, and Yong Li. 2022. Reinforcement Learning Enhances the Experts: Large-scale COVID-19 Vaccine Allocation with Multi-factor Contact Network. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3534678.3542679 [CORPUS].

[214] Nehal Hassan, Robert Slight, Kweku Bimpong, David W. Bates, Daniel Weiand, Akke Vellinga, Graham Morgan, and Sarah P. Slight. 2024. Systematic review to understand users perspectives on AI-enabled decision aids to inform shared decision making. *Nature Partner Journals Digital Medicine* (2024). doi:10.1038/s41746-024-01326-y .

[215] Joshua James Hatherley. 2020. Limits of trust in medical AI. *Journal of Medical Ethics* (2020). doi:10.2307/27197960 .

[216] Allyson I. Hauptman, Beau G. Schelble, Nathan J. McNeese, and Kapil Chalil Madathil. 2023. Adapt and overcome: perceptions of adaptive autonomous agents for human-ai teaming. *Computers in Human Behavior* (2023). doi:10.1016/j.chb.2022.107451 [CORPUS].

[217] Gaole He, Stefan Buijsman, and Ujwal Gadiraju. 2023. How Stated Accuracy of an AI System and Analogies to Explain Accuracy Affect Human Reliance on the System. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3610067 [CORPUS].

[218] Gaole He, Lucie Kuiper, and Ujwal Gadiraju. 2023. Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581025 [CORPUS].

[219] Xiangkun He and Chen Lv. 2023. Toward personalized decision making for autonomous vehicles: a constrained multi-objective reinforcement learning technique. *Transportation Research Part C: Emerging Technologies* (2023). doi:10.1016/j.trc.2023.104352 [CORPUS].

[220] Xiangkun He, Haohan Yang, Zhongxu Hu, and Chen Lv. 2023. Robust Lane Change Decision Making for Autonomous Vehicles: An Observation Adversarial Reinforcement Learning Approach. *IEEE Transactions on Intelligent Vehicles* (2023). doi:10.1109/TIV.2022.3165178 [CORPUS].

[221] Xin He, Xi Zheng, Huiyuan Ding, Yixuan Liu, and Hongling Zhu. 2024. Aicdss design guidelines and practice verification. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2023.2235882 [CORPUS].

[222] Oliver Hellum, Theis Ingerslev Jensen, Bryan T. Kelly, and Lasse Heje Pedersen. 2025. The Power of the Common Task Framework. SSRN. doi:10.2139/ssrn.5242901

[223] Patrick Hemmer, Max Schemmer, Lara Riefle, Nico Rosellen, Michael V¨ossing, and Niklas Kuehl. 2022. Factors that influence the adoption of human-ai collaboration in clinical decision-making. In *European Conference on Information Systems*. [CORPUS].

[224] Patrick Hemmer, Monika Westphal, Max Schemmer, Sebastian Vetter, Michael Vössing, and Gerhard Satzger. 2023. Human-AI Collaboration: The Effect of AI Delegation on Human Task Performance and Task Satisfaction. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3581641.3584052 [CORPUS].

[225] Christian Hendriksen. 2023. Artificial Intelligence for Supply Chain Management: Disruptive Innovation or Innovative Disruption? *Journal of Supply Chain Management* (2023). doi:10.1111/jscm.12304

[226] Erik Hermann, Gizem Yalcin Williams, and Stefano Puntoni. 2024. Deploying artificial intelligence in services to AID vulnerable consumers. *Journal of the Academy of Marketing Science* (2024). doi:10.1007/s11747-023-00986-8 [CORPUS].

[227] Sarita Herse, Jonathan Vitale, and Mary-Anne Williams. 2023. Using agent features to influence user trust, decision making and task outcome during human-agent collaboration. *International Journal of Human–Computer Interaction* (2023). doi:10.1080/10447318.2022.2150691 [CORPUS].

[228] Darin C. Hodges and A. F. Salam. 2018. Machine learning, analytics and strategic decision in the regulated energy industry. In *International Conference on Information Systems*. [CORPUS].

[229] Carl-Johan Hoel, Tommy Tram, and Jonas Sj¨oberg. 2020. Reinforcement Learning with Uncertainty Estimation for Tactical Decision-Making in Intersections. In *IEEE International Conference on Intelligent Transportation Systems*. doi:10.1109/ITSC45102.2020.9294407 [CORPUS].

[230] Lennart Hofeditz, S¨unje Clausen, Alexander Rieß, Milad Mirbabaie, and Stefan Stieglitz. 2022. Applying XAI to an AI-based system for candidate management to mitigate bias and discrimination in hiring. *Electronic Markets* (2022). doi:10.1007/s12525-022-00600-9 [CORPUS].

[231] Wayne Holmes and Ilkka Tuomi. 2022. State of the art and practice in AI in education. *European Journal of Education* 57, 4 (2022), 542–570.

[232] Kenneth Holstein, Maria De-Arteaga, Lakshmi Tumati, and Yanghuidi Cheng. 2023. Toward Supporting Perceptual Complementarity in Human-AI Collaboration via Reflection on Unobservables. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3579628 [CORPUS].

[233] Jiayi Hong, Ross Maciejewski, Alain Trubuil, and Tobias Isenberg. 2024. Visualizing and Comparing Machine Learning Predictions to Improve Human-AI Teaming on the Example of Cell Lineage. *IEEE Transactions on Visualization and Computer Graphics* (2024). doi:10.1109/TVCG.2023.3302308 [CORPUS].

[234] Wanshi Hong, Bin Wang, Mengqi Yao, Duncan Callaway, Larry Dale, and Can Huang. 2022. Data-driven power system optimal decision making strategy under wildfire events. In *Hawaii International Conference on System Sciences*. [CORPUS].

[235] Md Naimul Hoque and Klaus Mueller. 2022. Outcome-Explorer: A Causality Guided Interactive Visual Interface for Interpretable Algorithmic Decision Making. *IEEE Transactions on Visualization and Computer Graphics* (2022). doi:10.1109/TVCG.2021.3102051 [CORPUS].

[236] Min Hou, Le Wu, Enhong Chen, Zhi Li, Vincent W. Zheng, and Qi Liu. 2019. Explainable fashion recommendation: a semantic attribute region guided approach. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2019/650 [CORPUS].

[237] Yoyo Tsung-Yu Hou, Wen-Ying Lee, and Malte Jung. 2023. Should I Follow the Human, or Follow the Robot? — Robots in Power Can Have More Influence Than Humans on Decision-Making. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581066 [CORPUS].

[238] Brian Hu, Bill Ray, Alice Leung, Amy Summerville, David Joy, Christopher Funk, and Arslan Basharat. 2024. Language Models are Alignable Decision-Makers: Dataset and Application to the Medical Triage Domain. In *NAACL HLT (Industry Track)*. doi:10.18653/v1/2024.naacl-industry.18 [CORPUS].

[239] Mo Hu, Guanglu Zhang, Leah Chong, Jonathan Cagan, and Kosa Goucher-Lambert. 2024. How being outvoted by AI teammates impacts human-AI collaboration. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2345980 [CORPUS].

[240] Xiyang Hu, Yan Huang, Beibei Li, and Tian Lu. 2022. Credit Risk Modeling Without Sensitive Features: An Adversarial Deep Learning Model for Fairness and Profit. In *International Conference on Information Systems*. [CORPUS].

[241] Zhiyu Huang, Jingda Wu, and Chen Lv. 2022. Driving Behavior Modeling Using Naturalistic Human Driving Data With Inverse Reinforcement Learning. *IEEE Transactions on Intelligent Transportation Systems* (2022). doi:10.1109/TITS.2021.3088935 [CORPUS].

[242] Nicholas Huerta, Shiavax J. Rao, Ameesh Isath, Zhen Wang, Benjamin S. Glicksberg, and Chayakrit Krittanawong. 2024. The premise, promise, and perils of artificial intelligence in critical care cardiology. *Progress in Cardiovascular Diseases* (2024). doi:10.1016/j.pcad.2024.06.006 .

[243] Wasim Huleihel and Yehonathan Refael. 2024. Mathematical framework for online social media auditing. *Journal of Machine Learning Research* (2024). [CORPUS].

[244] Jessica Hullman, Alex Kale, and Jason Hartline. 2025. Underspecified Human Decision Experiments Considered Harmful. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Article 254, 15 pages. doi:10.1145/3706598.3714063

[245] Aziz Z. Huq. 2020. A right to a human decision. *Virginia Law Review* (2020). doi:10.2307/27074704 [CORPUS].

[246] Daphne Ippolito, Daniel Duckworth, Chris Callison-Burch, and Douglas Eck. 2020. Automatic Detection of Generated Text is Easiest when Humans are Fooled. In *Annual Meeting of the Association for Computational Linguistics*. doi:10.18653/v1/2020.acl-main.164 [CORPUS].

[247] Helmi Issa, Roy Dakroub, Hussein Lakkis, and Jad Jaber. 2024. Navigating the Decision-Making Landscape of AI in Risk Finance: Techno-Accountability Unveiled. *Risk Analysis: An International Journal* (2024). doi:10.1111/risa.14336 [CORPUS].

[248] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445385 [CORPUS].

[249] Johannes Jakubik, Jakob Sch¨offer, Vincent Hoge, Michael V¨ossing, and Niklas K¨uhl. 2023. An Empirical Evaluation of Predicted Outcomes as Explanations in Human-AI Decision-Making. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*. doi:10.1007/978-3-031-23618-1_24 [CORPUS].

[250] Prakash Jayakumar, Meredith G. Moore, Kenneth A. Furlough, Lauren M. Uhler, John P. Andrawis, Karl M. Koenig, Nazan Aksan, Paul J. Rathouz, and Kevin J. Bozic. 2021. Comparison of an artificial intelligence–enabled patient decision aid vs educational material on decision quality, shared decision-making, patient experience, and functional outcomes in adults with knee osteoarthritis: a randomized clinical trial. *JAMA Network Open* (2021). doi:10.1001/jamanetworkopen.2020.37107 [CORPUS].

[251] Jingru Jia, Zehua Yuan, Junhao Pan, Paul E McNamara, and Deming Chen. 2024. Decision-making behavior evaluation framework for llms under uncertain context. In *Advances in Neural Information Processing Systems*. [CORPUS].

[252] Yan Jia, John McDermid, Tom Lawton, and Ibrahim Habli. 2022. The Role of Explainability in Assuring Safety of Machine Learning in Healthcare. *IEEE Transactions on Emerging Topics in Computing* (2022). doi:10.1109/TETC.2022.3171314 [CORPUS].

[253] Jinglu Jiang, Surinder Kahai, and Ming Yang. 2022. Who needs explanation and when? Juggling explainable AI and user epistemic uncertainty. *International Journal of Human-Computer Studies* (2022). doi:10.1016/j.ijhcs.2022.102839 [CORPUS].

[254] Weina Jin, Xiaoxiao Li, Mostafa Fatehi, and Ghassan Hamarneh. 2023. Guidelines and Evaluation of Clinical Explainable AI in Medical Image Analysis. *Medical Image Analysis* (2023). doi:10.1016/j.media.2022.102684 [CORPUS].

[255] James Johnson. 2022. Delegating Strategic Decision-Making to Machines: Dr. Strangelove Redux? *Journal of Strategic Studies* (2022). doi:10.1080/01402390.2020.1759038 [CORPUS].

[256] Marina Johnson, Abdullah Albizri, Antoine Harfouche, and Samuel Fosso-Wamba. 2022. Integrating human knowledge into artificial intelligence for complex and ill-structured problems: informed artificial intelligence. *International Journal of Information Management* (2022). doi:10.1016/j.ijinfomgt.2022.102479 [CORPUS].

[257] Bryan D. Jones. 1999. BOUNDED RATIONALITY. *Annual Review of Political Science* 2, Volume 2, 1999 (1999), 297–321. doi:10.1146/annurev.polisci.2.1.297

[258] Shalmali Joshi, Sonali Parbhoo, and Finale Doshi-Velez. 2024. Learning-to-defer for sequential medical decision-making under uncertainty. *Transactions on Machine Learning Research* (2024). [CORPUS].

[259] Andreas Møller Jørgensen and Maria Appel Nissen. 2022. Making sense of decision support systems: rationales, translations and potentials for critical reflections on the reality of child protection. *Big Data & Society* (2022). doi:10.1177/20539517221125163 [CORPUS].

[260] Sergei V. Kalinin, Maxim Ziatdinov, Jacob Hinkle, Stephen Jesse, Ayana Ghosh, Kyle P. Kelley, Andrew R. Lupini, Bobby G. Sumpter, and Rama K. Vasudevan. 2021. Automated and Autonomous Experiments in Electron and Scanning Probe Microscopy. *ACS Nano* (2021). doi:10.1021/acsnano.1c02104 [CORPUS].

[261] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. "Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3517533 [CORPUS].

[262] Naveena Karusala, Sohini Upadhyay, Rajesh Veeraraghavan, and Krzysztof Z. Gajos. 2024. Understanding Contestability on the Margins: Implications for the Design of Algorithmic Decision-making in Public Services. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3641898 [CORPUS].

[263] Sonia K. Katyal. 2022. Democracy & distrust in an era of artificial intelligence. *Daedalus* (2022). doi:10.2307/48662045 [CORPUS].

[264] Harmanpreet Kaur, Cliff Lampe, and Walter S. Lasecki. 2020. Using affordances to improve AI support of social media posting decisions. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3377325.3377504 [CORPUS].

[265] Anna Kawakami, Amanda Coston, Hoda Heidari, Kenneth Holstein, and Haiyi Zhu. 2024. Studying Up Public Sector AI: How Networks of Power Relations Shape Agency Decisions Around AI Design and Use. *Proceedings of the ACM on Human-Computer Interaction* (2024). doi:10.1145/3686989 [CORPUS].

[266] Kenan Kaya, Carsten Gietzen, Robert Hahnfeldt, Maher Zoubi, Tilman Emrich, Moritz C. Halfmann, Malte Maria Sieren, Yannic Elser, Patrick Krumm, Jan M. Brendel, Konstantin Nikolaou, Nina Haag, Jan Borggrefe, Ricarda von Krüchten, Katharina Müller-Peltzer, Constantin Ehrengut, Timm Denecke, Andreas Hagendorff, Lukas Goertz, Roman J. Gertz, Alexander Christian Bunck, David Maintz, Thorsten Persigehl, Simon Lennartz, Julian A. Luetkens, Astha Jaiswal, Andra Iza Iuga, Lenhard Pennig, and Jonathan Kottlors. 2024. Generative Pre-trained Transformer 4 Analysis of Cardiovascular Magnetic Resonance Reports in Suspected Myocarditis: A Multicenter Study. *Journal of Cardiovascular Magnetic Resonance* (2024). doi:10.1016/j.jocmr.2024.101068 [CORPUS].

[267] Maxime Guillaume Kayser, Bayar Menzat, Cornelius Emde, Bogdan Alexandru Bercean, Alex Novak, Abdalá Trinidad Espinosa Morgado, Bartlomiej Papiez, Susanne Gaube, Thomas Lukasiewicz, and Oana-Maria Camburu. 2024. Fool Me Once? Contrasting Textual and Visual Explanations in a Clinical Decision-Support Setting. In *Empirical Methods in Natural Language Processing*. doi:10.18653/v1/2024.emnlp-main.1051 [CORPUS].

[268] Lauren E. Kelso, Jesse H. Grabman, David G. Dobolyi, and Chad S. Dodson. 2024. Does artificial intelligence (AI) assistance mitigate biased evaluations of eyewitness identifications? *Journal of Applied Research in Memory and Cognition* (2024). doi:10.1037/mac0000192 [CORPUS].

[269] Eoin M. Kenny, Elodie Ruelle, Anne Geoghegan, Laurence Shalloo, Micheál O'Leary, Michael O'Donovan, Mohammed Temraz, and Mark T. Keane. 2020. Bayesian case-exclusion and personalized explanations for sustainable dairy farming (extended abstract). In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2020/657 [CORPUS].

[270] Zine el Abidine Kherroubi, Samir Aknine, and Rebiha Bacha. 2022. Novel Decision-Making Strategy for Connected and Autonomous Vehicles in Highway On-Ramp Merging. *IEEE Transactions on Intelligent Transportation Systems* (2022). doi:10.1109/TITS.2021.3114983 [CORPUS].

[271] Naghmeh Khosrowabadi, Kai Hoberg, and Yun Shin Lee. 2024. Guiding supervisors in artificial intelligence-enabled forecasting: understanding the impacts of salience and detail on decision-making. *International Journal of Forecasting* (2024). doi:10.1016/j.ijforecast.2024.08.001 [CORPUS].

[272] Amirhossein Kiani, Bora Uyumazturk, Pranav Rajpurkar, Alex Wang, Rebecca Gao, Erik Jones, Yifan Yu, Curtis P. Langlotz, Robyn L. Ball, Thomas J. Montine, Brock A. Martin, Gerald J. Berry, Michael G. Ozawa, Florette K. Hazard, Ryanne A. Brown, Simon B. Chen, Mona Wood, Libby S. Allard, Lourdes Ylagan, Andrew Y. Ng, and Jeanne Shen. 2020. Impact of a deep learning assistant on the histopathologic classification of liver cancer. *Nature Partner Journals Digital Medicine* (2020). doi:10.1038/s41746-020-0232-8 [CORPUS].

[273] Buomsoo Kim, Jinsoo Park, and Jihae Suh. 2020. Transparency and accountability in AI decision support: explaining and visualizing convolutional neural networks for text information. *Decision Support Systems* (2020). doi:10.1016/j.dss.2020.113302 [CORPUS].

[274] Dajung Kim, Niko Vegt, Valentijn Visch, and Marina Bos-De Vos. 2024. How Much Decision Power Should (A)I Have?: Investigating Patients' Preferences Towards AI Autonomy in Healthcare Decision Making. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642883 [CORPUS].

[275] Jaeyoung Kim, Sihyeon Lee, Hyeon Jeon, Keon-Joo Lee, Hee-Joon Bae, Bohyoung Kim, and Jinwook Seo. 2024. PhenoFlow: A Human-LLM Driven Visual Analytics System for Exploring Large and Complex Stroke Datasets. *IEEE Transactions on Visualization and Computer Graphics* (2024). doi:10.1109/TVCG.2024.3456215 [CORPUS].

[276] Jeong Hyun Kim, Jungkeun Kim, Jooyoung Park, Changju Kim, Jihoon Jhang, and Brian King. 2023. When chatgpt gives incorrect answers: the impact of inaccurate information by generative ai on tourism decision-making. *Journal of Travel Research* (2023). doi:10.1177/00472875231212996 [CORPUS].

[277] Jee Young Kim, William Boag, Freya Gulamali, Alifia Hasan, Henry David Jeffry Hogg, Mark Lifson, Deirdre Mulligan, Manesh Patel, Inioluwa Deborah Raji, Ajai Sehgal, Keo Shaw, Danny Tobey, Alexandra Valladares, David Vidal, Suresh Balu, and Mark Sendak. 2023. Organizational Governance of Emerging Technologies: AI Adoption in Healthcare. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3593013.3594089 [CORPUS].

[278] Soonhee Kim, Kim Normann Andersen, and Jungwoo Lee. 2021. Platform government in the era of smart technology. *Public Administration Review* (2021). doi:10.1111/puar.13422 .

[279] Sunnie S. Y. Kim, Nicole Meister, Vikram V. Ramaswamy, Ruth Fong, and Olga Russakovsky. 2022. HIVE: Evaluating the Human Interpretability of Visual Explanations. In *European Conference on Computer Vision*. doi:10.1007/978-3-031-19775-8_17 [CORPUS].

[280] Taenyun Kim and Hayeon Song. 2023. Communicating the limitations of AI: The effect of message framing and ownership on trust in artificial intelligence. *International Journal of Human–Computer Interaction* (2023). doi:10.1080/10447318.2022.2049134 [CORPUS].

[281] Yubin Kim, Chanwoo Park, Hyewon Jeong, Yik Siu Chan, Xuhai Xu, Daniel McDuff, Hyeonhoon Lee, Marzyeh Ghassemi, Cynthia Breazeal, and Hae Won Park. 2024. Mdagents: An adaptive collaboration of LLMs for medical decision-making. In *Advances in Neural Information Processing Systems*. [CORPUS].

[282] Csaba Kindler, Stefan Elfwing, John Öhrvik, and Maziar Nikberg. 2023. A Deep Neural Network–Based Decision Support Tool for the Detection of Lymph Node Metastases in Colorectal Cancer Specimens. *Modern Pathology* (2023). doi:10.1016/j.modpat.2022.100015 [CORPUS].

[283] Artur Klingbeil, Cassandra Gr¨utzner, and Philipp Schreck. 2024. Trust and Reliance on AI — An Experimental Study on the Extent and Costs of Overreliance on AI. *Computers in Human Behavior* (2024). doi:10.1016/j.chb.2024.108352 [CORPUS].

[284] Daniel N. Kluttz and Deirdre K. Mulligan. 2019. Automated decision support technologies and the legal profession. *Berkeley Technology Law Journal* (2019). doi:10.2307/26954398 [CORPUS].

[285] Ioannis Koliousis, Abdulrahman Al-Surmi, and Mahdi Bashiri. 2024. Artificial intelligence and policy making; can small municipalities enable digital transformation? *International Journal of Production Economics* (2024). doi:10.1016/j.ijpe.2024.109324 [CORPUS].

[286] Flemming Kondrup, Thomas Jiralerspong, Elaine Lau, Nathan de Lara, Jacob Shkrob, My Duc Tran, Doina Precup, and Sumana Basu. 2024. Towards Safe Mechanical Ventilation Treatment Using Deep Offline Reinforcement Learning. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v37i13.26862 [CORPUS].

[287] Pascal D. K¨onig. 2022. Citizen conceptions of democracy and support for artificial intelligence in government and politics. *European Journal of Political Research* (2022). doi:10.1111/1475-6765.12570 [CORPUS].

[288] Yubo Kou and Xinning Gui. 2020. Mediating Community-AI Interaction through Situated Explanation: The Case of AI-Led Moderation. *Proceedings of the ACM on Human-Computer Interaction* (2020). doi:10.1145/3415173 [CORPUS].

[289] Dina Koutsikouri, Lena Hylving, Susanne Lindberg, and Jonna Bornemark. 2023. Seven elements of phronesis: a framework for understanding judgment in relation to automated decision-making. In *Hawaii International Conference on System Sciences*. [CORPUS].

[290] Mathias Kraus and Stefan Feuerriegel. 2017. Decision support from financial disclosures with deep neural networks and transfer learning. *Decision Support Systems* (2017). doi:10.1016/j.dss.2017.10.001 [CORPUS].

[291] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. Accountable algorithms. *University of Pennsylvania Law Review* (2017). doi:10.2307/26600576 [CORPUS].

[292] Marc Kuhn, Vanessa Reit, Maximilian Schwing, and Sarah Selinka. 2024. "Let the driver off the hook?" Moral decisions of autonomous cars and their impact on consumer well-being. *Transportation Research Part A: Policy and Practice* (2024). doi:10.1016/j.tra.2024.104224 [CORPUS].

[293] Sandeep Kumar, Hardik Arora, Tirthankar Ghosal, and Asif Ekbal. 2022. DeepASPeer: Towards an Aspect-level Sentiment Controllable Framework for Decision Prediction from Academic Peer Reviews. In *ACM/IEEE Joint Conference on Digital Libraries*. [CORPUS].

[294] Sandeep Kumar, Tirthankar Ghosal, Prabhat Kumar Bharti, and Asif Ekbal. 2021. Sharing is Caring! Joint Multitask Learning Helps Aspect-Category Extraction and Sentiment Detection in Scientific Peer Reviews. In *ACM/IEEE Joint Conference on Digital Libraries*. doi:10.1109/JCDL52503.2021.00081 [CORPUS].

[295] Matt Kusner, Chris Russell, Joshua Loftus, and Ricardo Silva. 2019. Making decisions that reduce discriminatory impacts. In *International Conference on Machine Learning (ICML)*. [CORPUS].

[296] Matt J. Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *Advances in Neural Information Processing Systems*. [CORPUS].

[297] Davide La Torre, Cinzia Colapinto, Ilaria Durosini, and Stefano Triberti. 2023. Team Formation for Human-Artificial Intelligence Collaboration in the Workplace: A Goal Programming Model to Foster Organizational Change. *IEEE Transactions on Engineering Management* (2023). doi:10.1109/TEM.2021.3077195 [CORPUS].

[298] Preethi Lahoti, Krishna P. Gummadi, and Gerhard Weikum. 2019. iFair: Learning Individually Fair Data Representations for Algorithmic Decision Making. In *IEEE International Conference on Data Engineering*. doi:10.1109/ICDE.2019.00121 [CORPUS].

[299] Vivian Lai, Samuel Carton, Rajat Bhatnagar, Q. Vera Liao, Yunfeng Zhang, and Chenhao Tan. 2022. Human-AI Collaboration via Conditional Delegation: A Case Study of Content Moderation. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3501999 [CORPUS].

[300] Vivian Lai, Chacha Chen, Alison Smith-Renner, Q. Vera Liao, and Chenhao Tan. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3593013.3594087

[301] Vivian Lai and Chenhao Tan. 2019. On Human Predictions with Explanations and Predictions of Machine Learning Models: A Case Study on Deception Detection. In *Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3287560.3287590 [CORPUS].

[302] Vivian Lai, Yiming Zhang, Chacha Chen, Q. Vera Liao, and Chenhao Tan. 2023. Selective Explanations: Leveraging Human Input to Align Explainable AI. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3610206 [CORPUS].

[303] Himabindu Lakkaraju, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2017. The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3097983.3098066 [CORPUS].

[304] Himabindu Lakkaraju and Cynthia Rudin. 2017. Learning Cost-Effective and Interpretable Treatment Regimes. In *International Conference on Artificial Intelligence and Statistics*. [CORPUS].

[305] Sophie Isabelle Lambert, Murielle Madi, Saša Sopka, Andrea Lenes, Hendrik Stange, Claus-Peter Buszello, and Astrid Stephan. 2023. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. *Nature Partner Journals Digital Medicine* (2023). doi:10.1038/s41746-023-00852-5 .

[306] Julius P. Landwehr, Niklas K¨uhl, Jannis Walk, and Mario Gn¨adig. 2023. Design knowledge for deep-learning-enabled image-based decision support systems. *Business & Information Systems Engineering* (2023). [CORPUS].

[307] Hannah S. Laqueur and Ryan W. Copus. 2024. An Algorithmic Assessment of Parole Decisions. *Journal of Quantitative Criminology* (2024). doi:10.1007/s10940-022-09563-8 [CORPUS].

[308] Daniel Laxar, Magdalena Eitenberger, Mathias Maleczek, Alexandra Kaider, Fabian Peter Hammerle, and Oliver Kimberger. 2023. The influence of explainable vs non-explainable clinical decision support systems on rapid triage

decisions: a mixed methods study. *BMC Medicine* (2023). doi:10.1186/s12916-023-03068-2 [CORPUS].

[309] Hansol Lee, René F. Kizilcec, and Thorsten Joachims. 2023. Evaluating a Learned Admission-Prediction Model as a Replacement for Standardized Tests in College Admissions. 195–203. doi:10.1145/3573051.3593382 [CORPUS].

[310] Kyootai Lee, Han-Gyun Woo, Wooje Cho, and Simon De Jong. 2022. When can AI reduce individuals' anchoring bias and enhance decision accuracy? Evidence from multiple longitudinal experiments. In *Hawaii International Conference on System Sciences*. [CORPUS].

[311] Kyungsik Lee, Hana Yoo, Sumin Shin, Wooyoung Kim, Yeonung Baek, Hyunjin Kang, Jaehyun Kim, and Kee-Eung Kim. 2024. A Submodular Optimization Approach to Accountable Loan Approval. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v38i21.30310 [CORPUS].

[312] Kyung Hwa Lee, Gwang Hyeon Choi, Jihye Yun, Jonggi Choi, Myung Ji Goh, Dong Hyun Sinn, Young Joo Jin, Minseok Albert Kim, Su Jong Yu, Sangmi Jang, Soon Kyu Lee, Jeong Won Jang, Jae Seung Lee, Do Young Kim, Young Youn Cho, Hyung Joon Kim, Sehwa Kim, Ji Hoon Kim, Namkug Kim, and Kang Mo Kim. 2024. Machine learning-based clinical decision support system for treatment recommendation and overall survival prediction of hepatocellular carcinoma: a multi-center study. *Nature Partner Journals Digital Medicine* (2024). doi:10.1038/s41746-023-00976-8 [CORPUS].

[313] Myeonghwa Lee, Seonho An, and Min-Soo Kim. 2024. PlanRAG: A Plan-then-Retrieval Augmented Generation for Generative Large Language Models as Decision Makers. In *Proceedings of the NAACL HLT*. doi:10.18653/v1/2024.naacl-long.364

[314] Min Hun Lee and Chong Jun Chew. 2023. Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3610218 [CORPUS].

[315] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2021. A Human-AI Collaborative Approach for Clinical Decision Making on Rehabilitation Assessment. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445472 [CORPUS].

[316] Min Hun Lee, Daniel P. Siewiorek, Asim Smailagic, Alexandre Bernardino, and Sergi Bermúdez i Badia. 2022. Towards Efficient Annotations for a Human-AI Collaborative, Clinical Decision Support System: A Case Study on Physical Stroke Rehabilitation Assessment. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3490099.3511112 [CORPUS].

[317] Min Kyung Lee and Katherine Rich. 2021. Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness around Healthcare AI and Cultural Mistrust. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445570 [CORPUS].

[318] Jingyuan Lei, Jizhuang Hui, Fengtian Chang, Salim Dassari, and Kai Ding. 2023. Reinforcement learning-based dynamic production-logistics-integrated tasks allocation in smart factories. *International Journal of Production Research* (2023). doi:10.1080/00207543.2022.2142314 [CORPUS].

[319] Christian Leibig, Moritz Brehmer, Stefan Bunk, Danalyn Byng, Katja Pinker, and Lale Umutlu. 2022. Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. *The Lancet Digital Health* (2022). doi:10.1016/S2589-7500(22)00070-X [CORPUS].

[320] Benedikt Leichtmann, Andreas Hinterreiter, Christina Humer, Marc Streit, and Martina Mara. 2024. Explainable artificial intelligence improves human decision-making: results from a mushroom picking experiment at a public art festival. *International Journal of Human–Computer Interaction* 40, 1 (2024), 1–23. doi:10.1080/10447318.2023.2221605 [CORPUS].

[321] Benedikt Leichtmann, Christina Humer, Andreas Hinterreiter, Marc Streit, and Martina Mara. 2023. Effects of explainable artificial intelligence on trust and human behavior in a high-risk decision task. *Computers in Human Behavior* (2023). doi:10.1016/j.chb.2022.107539 [CORPUS].

[322] Gabriel A. León, Erin K. Chiou, and Adam Wilkins. 2021. Accountability increases resource sharing: effects of accountability on human and AI system performance. *International Journal of Human–Computer Interaction* (2021). doi:10.1080/10447318.2020.1824695 [CORPUS].

[323] K. H. Leung, C. C. Luk, K. L. Choy, H. Y. Lam, and Carman K. M. Lee. 2019. A b2b flexible pricing decision support system for managing the request for quotation process under e-commerce business environment. *International Journal of Production Research* (2019). doi:10.1080/00207543.2019.1566674 [CORPUS].

[324] Michael Leyer and Sabrina Schneider. 2019. Me, you or AI? How do we feel about delegation. In *International Conference on Information Systems (ICIS)*. [CORPUS].

[325] Guofa Li, Siyan Lin, Shen Li, and Xingda Qu. 2022. Learning Automated Driving in Complex Intersection Scenarios Based on Camera Sensors: A Deep Reinforcement Learning Approach. *IEEE Sensors Journal* (2022). doi:10.1109/JSEN.2022.3146307 [CORPUS].

[326] Guofa Li, Yifan Yang, Shen Li, Xingda Qu, Nengchao Lyu, and Shengbo Eben Li. 2022. Decision making of autonomous vehicles in lane change scenarios: deep reinforcement learning approaches with risk awareness. *Transportation Research Part C: Emerging Technologies* (2022). doi:10.1016/j.trc.2021.103452 [CORPUS].

[327] Liangzhi Li, Kaoru Ota, and Mianxiong Dong. 2018. Humanlike Driving: Empirical Decision-Making System for Autonomous Vehicles. *IEEE Transactions on Vehicular Technology* (2018). doi:10.1109/TVT.2018.2822762 [CORPUS].

[328] Shaopeng Li and Teng Wu. 2022. Deep reinforcement learning-based decision support system for transportation infrastructure management under hurricane events. *Structural Safety* (2022). doi:10.1016/j.strusafe.2022.102254 [CORPUS].

[329] Tianlin Li, Qing Guo, Aishan Liu, Mengnan Du, Zhiming Li, and Yang Liu. 2023. Fairer: Fairness as decision rationale alignment. In *International Conference on Machine Learning (ICML)*. [CORPUS].

[330] Zhenhao Li, Tse-Hsun Chen, and Weiyi Shang. 2020. Where Shall We Log? Studying and Suggesting Logging Locations in Code Blocks. In *IEEE/ACM International Conference on Automated Software Engineering*. [CORPUS].

[331] Zhuoyan Li, Zhuoran Lu, and Ming Yin. 2023. Modeling Human Trust and Reliance in AI-Assisted Decision Making: A Markovian Approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v37i5.25748 [CORPUS].

[332] Zhuoyan Li, Zhuoran Lu, and Ming Yin. 2024. Decoding AI's Nudge: A Unified Framework to Predict Human Behavior in AI-Assisted Decision Making. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v38i9.28872 [CORPUS].

[333] Zhuoyan Li and Ming Yin. 2024. Utilizing human behavior modeling to manipulate explanations in AI-assisted decision making: the good, the bad, and the scary. In *Advances in Neural Information Processing Systems*. [CORPUS].

[334] Gabriel Lima, Nina Grgić-Hlača, and Meeyoung Cha. 2021. Human Perceptions on Moral Responsibility of AI: A Case Study in AI-Assisted Bail Decision-Making. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445260 [CORPUS].

[335] Raynaldio Limarga, Yang Song, Abhaya Nayak, David Rajaratnam, and Maurice Pagnucco. 2024. Formalisation and evaluation of properties for consequentialist machine ethics. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2024/49 [CORPUS].

[336] Jiaxin Lin, Wenhui Zhou, Hong Wang, Zhong Cao, Wenhao Yu, Chengxiang Zhao, Ding Zhao, Diange Yang, and Jun Li. 2022. Road Traffic Law Adaptive Decision-making for Self-Driving Vehicles. In *IEEE International Conference on Intelligent Transportation Systems*. doi:10.1109/ITSC55140.2022.9922208 [CORPUS].

[337] Fenglin Liu, Zheng Li, Hongjian Zhou, Qingyu Yin, Jingfeng Yang, Xianfeng Tang, Chen Luo, Ming Zeng, Haoming Jiang, Yifan Gao, Priyanka Nigam, Sreyashi Nag, Bing Yin, Yining Hua, Xuan Zhou, Omid Rohanian, Anshul Thakur, Lei Clifton, and David A. Clifton. 2024. Large Language Models Are Poor Clinical Decision-Makers: A Comprehensive Benchmark. In *Empirical Methods in Natural Language Processing*. doi:10.18653/v1/2024.emnlp-main.759 [CORPUS].

[338] Han Liu, Vivian Lai, and Chenhao Tan. 2021. Understanding the Effect of Out-of-distribution Examples and Interactive Explanations on Human-AI Decision Making. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3479552 [CORPUS].

[339] Jiaxin Liu, Hong Wang, Zhong Cao, Wenhao Yu, Chengxiang Zhao, Ding Zhao, Diange Yang, and Jun Li. 2023. Semantic Traffic Law Adaptive Decision-Making for Self-Driving Vehicles. *IEEE Transactions on Intelligent Transportation Systems* (2023). doi:10.1109/TITS.2023.3294579 [CORPUS].

[340] Michael Xieyang Liu, Aniket Kittur, and Brad A. Myers. 2022. Crystalline: Lowering the Cost for Developers to Collect and Organize Information for Decision Making. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3501968 [CORPUS].

[341] Qianyu Liu, Haoran Jiang, Zihao Pan, Qiushi Han, Zhenhui Peng, and Quan Li. 2024. BiasEye: A Bias-Aware Real-time Interactive Material Screening System for Impartial Candidate Assessment. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3640543.3645166 [CORPUS].

[342] Xueyi Liu, Qichao Zhang, Yinfeng Gao, and Zhongpu Xia. 2024. PnP: Integrated Prediction and Planning for Interactive Lane Change in Dense Traffic. In *Neural Information Processing*. doi:10.1007/978-981-99-8076-5_22 [CORPUS].

[343] Erping Long, Haotian Lin, Zhenzhen Liu, Xiaohang Wu, Liming Wang, Jiewei Jiang, Yingying An, Zhuoling Lin, Xiaoyan Li, Jingjing Chen, Jing Li, Qianzhong Cao, Dongni Wang, Xiyang Liu, Weirong Chen, and Yizhi Liu. 2017. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. *Nature Biomedical Engineering* (2017). doi:10.1038/s41551-016-0024 [CORPUS].

[344] Chiara Longoni and Luca Cian. 2022. Artificial intelligence in utilitarian vs. Hedonic contexts: the ""word-of-machine"" effect. *Journal of Marketing* (2022). doi:10.1177/0022242920957347 [CORPUS].

[345] Giorgia Lorenzini, Laura Arbelaez Ossa, David Martin Shaw, and Bernice Simone Elger. 2023. Artificial intelligence and the doctor–patient relationship expanding the paradigm of shared decision making. *Bioethics* (2023). doi:10.1111/bioe.13158 [CORPUS].

[346] Andrew Lowy, Devansh Gupta, and Meisam Razaviyayn. 2023. Stochastic differentially private and fair learning. In *International Conference on Learning Representations*. [CORPUS].

[347] Chao Lu, Hongliang Lu, Danni Chen, Haoyang Wang, Penghui Li, and Jianwei Gong. 2023. Human-like decision making for lane change based on the cognitive map and hierarchical reinforcement learning. *Transportation Research Part C: Emerging Technologies* (2023). doi:10.1016/j.trc.2023.104328 [CORPUS].

[348] Zhuoran Lu, Dakuo Wang, and Ming Yin. 2024. Does More Advice Help? The Effects of Second Opinions in AI-Assisted Decision Making. *Proceedings of the ACM on Human-Computer Interaction* (2024). doi:10.1145/3653708 [CORPUS].

[349] Zhuoran Lu and Ming Yin. 2021. Human Reliance on Machine Learning Models When Performance Feedback is Limited: Heuristics and Risks. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445562 [CORPUS].

[350] Brian Lubars and Chenhao Tan. 2019. Ask Not What AI Can Do, But What AI Should Do: Towards a Framework of Task Delegability. In *Advances in Neural Information Processing Systems*. [CORPUS].

[351] Adriano Lucieri, Muhammad Naseer Bajwa, Andreas Dengel, and Sheraz Ahmed. 2020. Explaining AI-Based Decision Support Systems Using Concept Localization Maps. In *Neural Information Processing*. doi:10.1007/978-3-030-63820-7_21 [CORPUS].

[352] Daniel J. Luckett, Eric B. Laber, Anna R. Kahkoska, David M. Maahs, Elizabeth Mayer-Davis, and Michael R. Kosorok. 2020. Estimating dynamic treatment regimes in mobile health using v-learning. *J. Amer. Statist. Assoc.* (2020). doi:10.1080/01621459.2018.1537919 [CORPUS].

[353] Jonas Ludwig, Paul-Michael Heineck, Marie-Theres Hess, Eleni Kremeti, Max Tauschhuber, Eric Hilgendorf, and Roland Deutsch. 2024. Inequality threat increases laypeople's, but not judges', acceptance of algorithmic decision making in court. *Law and Human Behavior* (2024). doi:10.1037/lhb0000577 [CORPUS].

[354] Fred C Lunenburg. 2010. The decision making process.. In *National Forum of Educational Administration & Supervision Journal*, Vol. 27.

[355] Pin Lv, Jinlei Han, Jiangtian Nie, Yang Zhang, Jia Xu, Chao Cai, and Zhe Chen. 2023. Cooperative Decision-Making of Connected and Autonomous Vehicles in an Emergency. *IEEE Transactions on Vehicular Technology* (2023). doi:10.1109/TVT.2022.3211884 [CORPUS].

[356] Henrietta Lyons, Eduardo Velloso, and Tim Miller. 2021. Conceptualising Contestability: Perspectives on Contesting Algorithmic Decisions. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3449180 [CORPUS].

[357] Karol Lina López, Christian Gagné, and Marc-André Gardner. 2019. Demand-Side Management Using Deep Learning for Smart Charging of Electric Vehicles. *IEEE Transactions on Smart Grid* (2019). doi:10.1109/TSG.2018.2808247 [CORPUS].

[358] Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2023. Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581058 [CORPUS].

[359] Shuai Ma, Xinru Wang, Ying Lei, Chuhan Shi, Ming Yin, and Xiaojuan Ma. 2024. "Are You Really Sure?" Understanding the Effects of Human Self-Confidence Calibration in AI-Assisted Decision Making. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642671 [CORPUS].

[360] I. Scott MacKenzie. 1992. Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human–Computer Interaction* 7, 1 (1992), 91–139.

[361] Geetha Mahadevaiah, Prasad RV, Inigo Bermejo, David Jaffray, Andre Dekker, and Leonard Wee. 2020. Artificial intelligence-based clinical decision support in modern medical physics: selection, acceptance, commissioning, and quality assurance. *Medical Physics* (2020). doi:10.1002/mp.13562 [CORPUS].

[362] Syed Hasan Amin Mahmood, Zhuoran Lu, and Ming Yin. 2024. Designing behavior-aware AI to improve the human-AI team performance in AI-assisted decision making. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2024/344 [CORPUS].

[363] Marian Marchal, Merel Scholman, Frances Yung, and Vera Demberg. 2022. Establishing Annotation Quality in Multi-label Annotations. In *Proceedings of the International Conference on Computational Linguistics*, Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na (Eds.). International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 3659–3668. https://aclanthology.org/2022.coling-1.322/

[364] Michael A. Marchetti, Emily A. Cowen, Nicholas R. Kurtansky, Jochen Weber, Megan Dauscher, Jennifer DeFazio, Liang Deng, Stephen W. Dusza, Helen Haliasos, Allan C. Halpern, Sharif Hosein, Zaeem H. Nazir, Ashfaq A. Marghoob, Elizabeth A. Quigley, Trina Salvador, and Veronica M. Rotemberg. 2023. Prospective validation of dermoscopy-based open-source artificial intelligence for melanoma diagnosis (PROVE-AI study). *Nature Partner Journals Digital Medicine* (2023). doi:10.1038/s41746-023-00872-1 [CORPUS].

[365] Frank Marcinkowski, Kimon Kieslich, Christopher Starke, and Marco L"unich. 2020. Implications of AI (un-)fairness in higher education admissions: the effects of perceived AI (un-)fairness on exit, voice and organizational reputation. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3351095.3372867 [CORPUS].

[366] Alex Mari and René Algesheimer. 2021. The role of trusting beliefs in voice assistants during voice shopping. In *Hawaii International Conference on System Sciences*. [CORPUS].

[367] Alex Mari, Andreina Mandelli, and René Algesheimer. 2024. Empathic voice assistants: enhancing consumer responses in voice commerce. *Journal of Business Research* (2024). doi:10.1016/j.jbusres.2024.114566 [CORPUS].

[368] Deniz Marti, Anjila Budathoki, Yi Ding, Gale Lucas, and David Nelson. 2024. How does acknowledging users' preferences impact AI's ability to make conflicting recommendations? *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2426035 [CORPUS].

[369] Guillaume Martin and Rapha"el Oger. 2022. A Reinforcement Learning Powered Digital Twin to Support Supply Chain Decisions. In *Hawaii International Conference on System Sciences*. [CORPUS].

[370] Kirsten Martin and Ari Waldman. 2023. Are Algorithmic Decisions Legitimate? The Effect of Process and Outcomes on Perceptions of Legitimacy of AI Decisions. *Journal of Business Ethics* (2023). doi:10.1007/s10551-021-05032-7 [CORPUS].

[371] Marina Martín and José A. Macías. 2023. A Supporting Tool for Enhancing User's Mental Model Elicitation and Decision-Making in User Experience Research. *International Journal of Human–Computer Interaction* (2023). doi:10.1080/10447318.2022.2041885 [CORPUS].

[372] Jorge Martín-Pérez, Kiril Antevski, Andres Garcia-Saavedra, Xi Li, and Carlos J. Bernardos. 2021. DQN Dynamic Pricing and Revenue Driven Service Federation Strategy. *IEEE Transactions on Network and Service Management* (2021). doi:10.1109/TNSM.2021.3117589 [CORPUS].

[373] Marc Maurer, Tolga Buz, Christian Dremel, and Gerard de Melo. 2024. Design and Evaluation of an AI-Augmented Screening System for Venture Capitalists. In *European Conference on Information Systems*. [CORPUS].

[374] Melissa Mccradden, Oluwadara Odusi, Shalmali Joshi, Ismail Akrout, Kagiso Ndlovu, Ben Glocker, Gabriel Maicas, Xiaoxuan Liu, Mjaye Mazwi, Tee Garnett, Lauren Oakden-Rayner, Myrtede Alfred, Irvine Sihlahla, Oswa Shafei, and Anna Goldenberg. 2023. What's fair is... fair? Presenting JustEFAB, an ethical framework for operationalizing medical ethics and social justice in the integration of clinical machine learning: JustEFAB. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3593013.3594096 [CORPUS].

[375] Duncan C. McElfresh, Lok Chan, Kenzie Doyle, Walter Sinnott-Armstrong, Vincent Conitzer, Jana Schaich Borg, and John P. Dickerson. 2021. Indecision Modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v35i7.16746 [CORPUS].

[376] Reid McIlroy-Young, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. 2020. Aligning Superhuman AI with Human Behavior: Chess as a Model System. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3394486.3403219 [CORPUS].

[377] Geoffrey Mead and Barbara Barbosa Neves. 2023. Contested delegation: understanding critical public responses to algorithmic decision-making in the UK and Australia. *The Sociological Review* (2023). doi:10.1177/00380261221105380 [CORPUS].

[378] Esther D. Meenken, Christopher M. Triggs, Hamish E. Brown, Sarah Sinton, Jeremy R. Bryant, Alasdair D.L. Noble, Martin Espig, Mostafa Sharifi, and David M. Wheeler. 2021. Bayesian hybrid analytics for uncertainty analysis and real-time crop management. *Agronomy Journal* 113, 3 (2021), 2491–2505. doi:10.1002/agj2.20659 [CORPUS].

[379] Hugh Mehan. 1983. The role of language and the language of role in institutional decision making. *Language in Society* 12, 2 (1983), 187–211. doi:10.1017/S0047404500009805

[380] Tiago Melo, Altigran S. Silva, Edleno S. Moura, and Pável Calado. 2019. Contender: Leveraging User Opinions for Purchase Decision-Making. In *European Conference on Information Retrieval*. doi:10.1007/978-3-030-15719-7_30 [CORPUS].

[381] Fanlin Meng, Xiao-Jun Zeng, Yan Zhang, Chris J. Dent, and Dunwei Gong. 2018. An integrated optimization + learning approach to optimal dynamic pricing for the retailer with multi-type customers in smart grids. *Information Sciences* (2018). doi:10.1016/j.ins.2018.03.039 [CORPUS].

[382] Sarah Mertens, Joachim Krois, Anselmo Garcia Cantu, Lubaina T. Arsiwala, and Falk Schwendicke. 2021. Artificial intelligence for caries detection: randomized trial. *Journal of Dentistry* (2021). doi:10.1016/j.jdent.2021.103849 [CORPUS].

[383] Lewis H. Mervin, Simon Johansson, Elizaveta Semenova, Kathryn A. Giblin, and Ola Engkvist. 2021. Uncertainty quantification in drug design. *Drug Discovery Today* (2021). doi:10.1016/j.drudis.2020.11.027 .

[384] Bonan Min, Benjamin Rozonoyer, Haoling Qiu, Alexander Zamanian, Nianwen Xue, and Jessica MacBride. 2021. ExcavatorCovid: Extracting Events and Relations from Text Corpora for Temporal and Causal Analysis for COVID-19. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. doi:10.18653/v1/2021.emnlp-demo.8 [CORPUS].

[385] Shalini Misra, Benjamin Katz, Patrick Roberts, Mackenzie Carney, and Isabel Valdivia. 2024. Toward a person-environment fit framework for artificial intelligence implementation in the public sector. *Government Information Quarterly*

(2024). doi:10.1016/j.giq.2024.101962 [CORPUS].
[386] Miguel A. Mohedano-Munoz, Cristina Soguero-Ruiz, Inmaculada Mora-Jiménez, Manuel Rubio-Sánchez, Joaquín Álvarez Rodríguez, and Alberto Sanchez. 2023. A streaming data visualization framework for supporting decision-making in the intensive care unit. *Expert Systems with Applications* (2023). doi:10.1016/j.eswa.2023.120252 [CORPUS].
[387] Karishma Mohiuddin, Mirza Ariful Alam, Mirza Mohtashim Alam, Pascal Welke, Michael Martin, Jens Lehmann, and Sahar Vahdati. 2023. Retention is All You Need. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. doi:10.1145/3583780.3615497 [CORPUS].
[388] Katelyn Morrison, Donghoon Shin, Kenneth Holstein, and Adam Perer. 2023. Evaluating the Impact of Human Explanation Strategies on Human-AI Visual Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* (2023). doi:10.1145/3579481 [CORPUS].
[389] Ece Çiğdem Mutlu, Niloofar Yousefi, and Ozlem Ozmen Garibay. 2022. Contrastive Counterfactual Fairness in Algorithmic Decision-Making. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534143 [CORPUS].
[390] NA. 2024. Designing Skill-Compatible AI: Methodologies and Frameworks in Chess. In *International Conference on Learning Representations*. [CORPUS].
[391] N/A. 2024. Dilu: a knowledge-driven approach to autonomous driving with large language models. In *International Conference on Learning Representations*. [CORPUS].
[392] NA. 2024. Empirical likelihood for fair classification. In *International Conference on Learning Representations*. [CORPUS].
[393] NA. 2024. Hazard Challenge: Embodied Decision Making in Dynamically Changing Environments. In *International Conference on Learning Representations*. [CORPUS].
[394] Razieh Nabi, Daniel Malinsky, and Ilya Shpitser. 2019. Learning optimal fair policies. In *International Conference on Machine Learning (ICML)*. [CORPUS].
[395] Myura Nagendran, Paul Festor, Matthieu Komorowski, Anthony C. Gordon, and Aldo A. Faisal. 2023. Quantifying the impact of AI recommendations with explanations on prescription decision making. *Nature Partner Journals Digital Medicine* (2023). doi:10.1038/s41746-023-00955-z [CORPUS].
[396] Hsiang Sing Naik, Jiaoping Zhang, Alec Lofquist, Teshale Assefa, Soumik Sarkar, David Ackerman, Arti Singh, Asheesh K. Singh, and Baskar Ganapathysubramanian. 2017. A real-time phenotyping framework using machine learning for plant stress severity rating in soybean. *Plant Methods* (2017). doi:10.1186/s13007-017-0173-7 [CORPUS].
[397] Mohammad Naiseh, Dena Al-Thani, Nan Jiang, and Raian Ali. 2021. Explainable recommendation: when design meets trust calibration. *World Wide Web Journal* (2021). doi:10.1007/s11280-021-00916-0 [CORPUS].
[398] Mohammad Naiseh, Dena Al-Thani, Nan Jiang, and Raian Ali. 2023. How the different explanation classes impact trust calibration: the case of clinical decision support systems. *International Journal of Human-Computer Studies* (2023). doi:10.1016/j.ijhcs.2022.102941 [CORPUS].
[399] Kazuki Nakamura, Ryosuke Kojima, Eiichiro Uchino, Koh Ono, Motoko Yanagita, Koichi Murashita, Ken Itoh, Shigeyuki Nakaji, and Yasushi Okuno. 2021. Health improvement framework for actionable treatment planning using a surrogate Bayesian model. *Nature Communications* (2021). doi:10.1038/s41467-021-23319-1 [CORPUS].
[400] Mila Nambiar, Yong Mong Bee, Yu En Chan, Ivan Ho Mien, Feri Guretno, David Carmody, Phong Ching Lee, Sing Yi Chia, Nur Nasyitah Mohamed Salim, and Pavitra Krishnaswamy. 2024. A drug mix and dose decision algorithm for individualized type 2 diabetes management. *Nature Partner Journals Digital Medicine* (2024). doi:10.1038/s41746-024-01230-5 [CORPUS].
[401] Hongseok Namkoong, Ramtin Keramati, Steve Yadlowsky, and Emma Brunskill. 2020. Off-policy policy evaluation for sequential decisions under unobserved confounding. In *Advances in Neural Information Processing Systems*. [CORPUS].
[402] Steffen Nauhaus, Johannes Luger, and Sebastian Raisch. 2021. Strategic decision making in the digital age: expert sentiment and corporate capital allocation. *Journal of Management Studies* (2021). doi:10.1111/joms.12742 [CORPUS].
[403] Caroline A. Nelson, Lourdes Maria Pérez-Chada, Andrew Creadore, Sara Jiayang Li, Kelly Lo, Priya Manjaly, Ashley Bahareh Pournamdari, Elizabeth Tkachenko, John S. Barbieri, Justin M. Ko, Alka V. Menon, Rebecca Ivy Hartman, and Arash Mostaghimi. 2020. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. *Journal of the American Medical Association Dermatology* (2020). doi:10.1001/jamadermatol.2019.5014 [CORPUS].
[404] Christopher Nemeth, Josh Blomberg, Christopher Argenta, Maria L. Serio-Melvin, Jose Salinas, and Jeremy Pamplin. 2016. Revealing icu cognitive work through naturalistic decision-making methods. *Journal of Cognitive Engineering and Decision Making* (2016). doi:10.1177/1555343416664845 [CORPUS].
[405] Evangelos Niforatos, Adam Palma, Roman Gluszny, Athanasios Vourvopoulos, and Fotis Liarokapis. 2020. Would you do it?: Enacting Moral Dilemmas in Virtual Reality for Understanding Ethical Decision-Making. In *Conference on Human Factors in Computing Systems*. doi:10.1145/3313831.3376788 [CORPUS].
[406] Hamed Nilforoshan, Johann D Gaebler, Ravi Shroff, and Sharad Goel. 2022. Causal conceptions of fairness and their consequences. In *International Conference on Machine Learning (ICML)*. [CORPUS].
[407] Revital Nimri, Tadej Battelino, Lori M. Laffel, Robert H. Slover, Desmond Schatz, Stuart A. Weinzimer, Klemen Dovc, Thomas Danne, Moshe Phillip, and NextDREAM Consortium. 2020. Insulin dose optimization using an automated artificial intelligence-based decision support system in youths with type 1 diabetes. *Nature Medicine* (2020). doi:10.1038/s41591-020-1045-7 [CORPUS].
[408] Rohit Nishant, Dirk Schneckenberg, and MN Ravishankar. 2024. The formal rationality of artificial intelligence-based algorithms and the problem of bias. *Journal of Information Technology* (2024). [CORPUS].
[409] Ritesh Noothigattu, Snehalkumar Gaikwad, Edmond Awad, Sohan Dsouza, Iyad Rahwan, Pradeep Ravikumar, and Ariel Procaccia. 2018. A Voting-Based System for Ethical Decision Making. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v32i1.11512 [CORPUS].
[410] Alejandro Noriega-Campero, Michiel A. Bakker, Bernardo Garcia-Bulle, and Alex 'Sandy' Pentland. 2019. Active Fairness in Algorithmic Decision Making. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3306618.3314277 [CORPUS].
[411] Mahsan Nourani, Chiradeep Roy, Jeremy E Block, Donald R Honeycutt, Tahrima Rahman, Eric Ragan, and Vibhav Gogate. 2021. Anchoring Bias Affects Mental Model Formation and User Reliance in Explainable AI Systems. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3397481.3450639 [CORPUS].
[412] Anne-Marie Nussberger, Lan Luo, L. Elisa Celis, and M. J. Crockett. 2022. Public attitudes value interpretability but prioritize accuracy in Artificial Intelligence. *Nature Communications* (2022). doi:10.1038/s41467-022-33417-3 [CORPUS].
[413] Yujin Oh, Go Eun Bae, Kyung-Hee Kim, Min-Kyung Yeo, and Jong Chul Ye. 2023. Multi-Scale Hybrid Vision Transformer for Learning Gastric Histology: AI-Based Decision Support System for Gastric Cancer Treatment. *IEEE Journal of Biomedical and Health Informatics* (2023). doi:10.1109/JBHI.2023.3276778 [CORPUS].
[414] Harley Oliff, Ying Liu, Maneesh Kumar, Michael Williams, and Michael Ryan. 2020. Reinforcement learning for facilitating human-robot-interaction in manufacturing. *Journal of Manufacturing Systems* (2020). doi:10.1016/j.jmsy.2020.06.018 [CORPUS].
[415] Başak Oral, Pierre Dragicevic, Alexandru Telea, and Evanthia Dimara. 2024. Decoupling Judgment and Decision Making: A Tale of Two Tails. *IEEE Transactions on Visualization and Computer Graphics* 30, 10 (2024), 6928–6940. doi:10.1109/TVCG.2023.3346640
[416] Cassandra Overney, Belén Saldías, Dimitra Dimitrakopoulou, and Deb Roy. 2024. SenseMate: An Accessible and Beginner-Friendly Human-AI Platform for Qualitative Data Analysis. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3640543.3645194 [CORPUS].
[417] Daniel Owen. 2015. Collaborative Decision Making. *Decision Analysis* 12, 1 (2015), 29–45. doi:10.1287/deca.2014.0307
[418] Aldo Pacchiano, Shaun Singh, Edward Chou, Alex Berg, and Jakob Foerster. 2021. Neural pseudo-label optimism for the bank loan problem. In *Advances in Neural Information Processing Systems*. [CORPUS].
[419] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372 (2021). doi:10.1136/bmj.n71
[420] Evgeny Palchevsky, Vyacheslav Antonov, Rustem Radomirovich Enikeev, and Tim Breikin. 2023. A system based on an artificial neural network of the second generation for decision support in especially significant situations. *Journal of Hydrology* (2023). doi:10.1016/j.jhydrol.2022.128844 [CORPUS].
[421] Shuyi Pan and Yi Mou. 2024. Team up with AI or human? Investigating candidates' self-categorization as fluidity and ingroup-serving attribution when judged by a human–AI hybrid jury. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2408511 [CORPUS].
[422] Ravi Pandya, Sandy H. Huang, Dylan Hadfield-Menell, and Anca D. Dragan. 2019. Human-AI Learning Performance in Multi-Armed Bandits. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3306618.3314245 [CORPUS].
[423] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. 2023. Co-design of Human-centered, Explainable AI for Clinical Decision Support. *ACM Transactions on Interactive Intelligent Systems* (2023). doi:10.1145/3587271 [CORPUS].
[424] Cecilia Panigutti, Andrea Beretta, Fosca Giannotti, and Dino Pedreschi. 2022. Understanding the Impact of Explanations on Advice-Taking: A User Study for AI-Based Clinical Decision Support Systems. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3502104 [CORPUS].

[425] Cecilia Panigutti, Alan Perotti, André Panisson, Paolo Bajardi, and Dino Pedreschi. 2021. Fairlens: Auditing Black-box Clinical Decision Support Systems. *Information Processing & Management* (2021). doi:10.1016/j.ipm.2021.102657 [CORPUS].

[426] Andrea Papenmeier, Dagmar Kern, Gwenn Englebienne, and Christin Seifert. 2022. It's Complicated: The Relationship between User Trust, Model Accuracy and Explanations in AI. *ACM Transactions on Computer-Human Interaction* (2022). doi:10.1145/3495013 [CORPUS].

[427] Saumya Pareek, Eduardo Velloso, and Jorge Goncalves. 2024. Trust Development and Repair in AI-Assisted Decision-Making during Complementary Expertise. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3630106.3658924 [CORPUS].

[428] Eun Hee Park, Karl Werder, Lan Cao, and Balasubramaniam Ramesh. 2022. Why do family members reject AI in health care? Competing effects of emotions. *Journal of Management Information Systems* (2022). doi:10.1080/07421222.2022.2096550 [CORPUS].

[429] Hyanghee Park, Daehwan Ahn, Kartik Hosanagar, and Joonhwan Lee. 2021. Human-AI Interaction in Human Resource Management: Understanding Why Employees Resist Algorithmic Evaluation at Workplaces and How to Mitigate Burdens. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445304 [CORPUS].

[430] Kidon Park, Hong-Gyu Jung, Tae-San Eom, and Seong-Whan Lee. 2024. Uncertainty-Aware Portfolio Management With Risk-Sensitive Multiagent Network. *IEEE Transactions on Neural Networks and Learning Systems* (2024). doi:10.1109/TNNLS.2022.3174642 [CORPUS].

[431] Mario Passalacqua, Robert Pellerin, Esma Yahia, Florian Magnani, Frédéric Rosin, Laurent Joblot, and Pierre-Majorique Léger. 2025. Practice with less AI makes perfect: Partially automated AI during training leads to better worker motivation, engagement, and skill acquisition. *International Journal of Human–Computer Interaction* (2025). doi:10.1080/10447318.2024.2319914 [CORPUS].

[432] Elisabeth Paté-Cornell. 2024. Preferences in AI Algorithms: The Need for Relevant Risk Attitudes in Automated Decisions Under Uncertainties. *Risk Analysis: An International Journal* (2024). doi:10.1111/risa.14268 [CORPUS].

[433] N. Peiffer-Smadja, T.M. Rawson, R. Ahmad, A. Buchard, P. Georgiou, F.-X. Lescure, G. Birgand, and A.H. Holmes. 2020. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clinical Microbiology and Infection* (2020). doi:10.1016/j.cmi.2019.09.009 .

[434] Andi Peng, Besmira Nushi, Emre Kiciman, Kori Inkpen, and Ece Kamar. 2022. Investigations of Performance and Bias in Human-AI Teamwork in Hiring. In *AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v36i11.21468 [CORPUS].

[435] Dana Pessach, Gonen Singer, Dan Avrahami, Hila Chalutz Ben-Gal, Erez Shmueli, and Irad Ben-Gal. 2020. Employees recruitment: a prescriptive analytics approach via machine learning and mathematical programming. *Decision Support Systems* (2020). doi:10.1016/j.dss.2020.113290 [CORPUS].

[436] Florian Pethig and Julia Kroenung. 2023. Biased Humans, (Un)Biased Algorithms? *Journal of Business Ethics* (2023). doi:10.1007/s10551-022-05071-8 [CORPUS].

[437] Daniele Petrone, Neofytos Rodosthenous, and Vito Latora. 2022. An AI approach for managing financial systemic risk via bank bailouts by taxpayers. *Nature Communications* (2022). doi:10.1038/s41467-022-34102-1 [CORPUS].

[438] Evangelos Pournaras. 2020. Proof of witness presence: blockchain consensus for augmented democracy in smart cities. *J. Parallel and Distrib. Comput.* (2020). doi:10.1016/j.jpdc.2020.06.015 [CORPUS].

[439] Snehal Prabhudesai, Leyao Yang, Sumit Asthana, Xun Huan, Q. Vera Liao, and Nikola Banovic. 2023. Understanding Uncertainty: How Lay Decision-makers Perceive and Interpret Uncertainty in Human-AI Decision Making. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3581641.3584033 [CORPUS].

[440] Ashish Viswanath Prakash and Saini Das. 2021. Medical practitioner's adoption of intelligent clinical diagnostic decision support systems: a mixed-methods study. *Information & Management* (2021). doi:10.1016/j.im.2021.103524 [CORPUS].

[441] Paul Prinsloo, Mohammad Khalil, and Sharon Slade. 2024. Vulnerable student digital well-being in AI-powered educational decision support systems (AI-EDSS) in higher education. *British Journal of Educational Technology* (2024). doi:10.1111/bjet.13508 [CORPUS].

[442] Paul Prinsloo, Sharon Slade, and Mohammad Khalil. 2023. At the intersection of human and algorithmic decision-making in distributed learning. *Journal of Research on Technology in Education* (2023). doi:10.1080/15391523.2022.2121343 [CORPUS].

[443] Paolo Priore, Borja Ponte, Rafael Rosillo, and David de la Fuente. 2019. Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. *International Journal of Production Research* (2019). doi:10.1080/00207543.2018.1552369 [CORPUS].

[444] Bhagyashree Puranik, Upamanyu Madhow, and Ramtin Pedarsani. 2022. A Dynamic Decision-Making Framework Promoting Long-Term Fairness. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534127 [CORPUS].

[445] David Quarfoot and Richard A. Levine. 2016. How Robust Are Multirater Interrater Reliability Indices to Changes in Frequency Distribution? *The American Statistician* 70, 4 (2016), 373–384. doi:10.1080/00031305.2016.1141708

[446] Piyya Muhammad Rafi-Ul-Shan, Mahdi Bashiri, Muhammad Mustafa Kamal, Sachin Kumar Mangla, and Benny Tjahjono. 2024. An Analysis of Fuzzy Group Decision Making to Adopt Emerging Technologies for Fashion Supply Chain Risk Management. *IEEE Transactions on Engineering Management* (2024). doi:10.1109/TEM.2024.3354845 [CORPUS].

[447] Orit Raphaeli and Pierre Singer. 2021. Towards personalized nutritional treatment for malnutrition using machine learning-based screening tools. *Clinical Nutrition* (2021). doi:10.1016/j.clnu.2021.08.013 [CORPUS].

[448] Charvi Rastogi, Yunfeng Zhang, Dennis Wei, Kush R. Varshney, Amit Dhurandhar, and Richard Tomsett. 2022. Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction* (2022). doi:10.1145/3512930 [CORPUS].

[449] Omer Reingold, Judy Hanwen Shen, and Aditi Talati. 2024. Dissenting Explanations: Leveraging Disagreement to Reduce Model Overreliance. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v38i19.30151 [CORPUS].

[450] Michael Allan Ribers and Hannes Ullrich. 2024. Complementarities between algorithmic and human decision-making: The case of antibiotic prescribing. *Quantitative Marketing and Economics* (2024). doi:10.1007/s11129-024-09284-1 [CORPUS].

[451] René Riedl. 2022. Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets* (2022). doi:10.1007/s12525-022-00594-4 .

[452] Juan-Pablo Rivera, Gabriel Mukobi, Anka Reuel, Max Lamparth, Chandler Smith, and Jacquelyn Schneider. 2024. Escalation Risks from Language Models in Military and Diplomatic Decision-Making. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3630106.3658942 [CORPUS].

[453] Stefano Giovanni Rizzo, Yixian Chen, Linsey Pang, Ji Lucas, Zoi Kaoudi, Jorge Quiane, and Sanjay Chawla. 2020. Prescriptive Learning for Air-Cargo Revenue Management. In *IEEE International Conference on Data Mining*. doi:10.1109/ICDM50108.2020.00055 [CORPUS].

[454] Vincent Robbemond, Oana Inel, and Ujwal Gadiraju. 2022. Understanding the Role of Explanation Modality in AI-assisted Decision-making. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. doi:10.1145/3503252.3531311 [CORPUS].

[455] Waymond Rodgers, James M. Murray, Abraham Stefanidis, William Y. Degbey, and Shlomo Y. Tarba. 2023. An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes. *Human Resource Management Review* (2023). doi:10.1016/j.hrmr.2022.100925 [CORPUS].

[456] Simon P Rowland, J. Edward Fitzgerald, Matthew Lungren, Elizabeth (Hsieh) Lee, Zach Harned, and Alison H. McGregor. 2022. Digital health technology-specific risks for medical malpractice liability. *Nature Partner Journals Digital Medicine* (2022). doi:10.1038/s41746-022-00698-3 .

[457] Jože M. Rožanec, Inna Novalija, Patrik Zajec, Klemen Kenda, Hooman Tavakoli Ghinani, Sungho Suh, Entso Veliou, Dimitrios Papamartzivanos, Thanassis Giannetsos, Sofia Anna Menesidou, Ruben Alonso, Nino Cauli, Antonello Meloni, Diego Reforgiato Recupero, Dimosthenis Kyriazis, Georgios Sofianidis, Spyros Theodoropoulos, Blaž Fortuna, Dunja Mladenić, and John Soldatos. 2023. Human-centric artificial intelligence architecture for industry 5.0 applications. *International Journal of Production Research* (2023). doi:10.1080/00207543.2022.2138611 [CORPUS].

[458] Gloire Rubambiza, Fernando Romero Galvan, Ryan Pavlick, Hakim Weatherspoon, and Kaitlin M. Gold. 2023. Toward cloud-native, machine learning based detection of crop disease with imaging spectroscopy. *Journal of Geophysical Research: Biogeosciences* (2023). doi:10.1029/2022JG007342 [CORPUS].

[459] Brian K. Russell, Josh McGeown, and Bettina L. Beard. 2023. Developing AI Enabled Sensors and Decision Support for Military Operators in the Field. *Journal of Science and Medicine in Sport* (2023). doi:10.1016/j.jsams.2023.03.001 [CORPUS].

[460] Patrik Sabol, Peter Sinčák, Pitoyo Hartono, Pavel Kočan, Zuzana Benetinová, Alžbeta Blichárová, Ľudmila Verbóová, Erika Štammová, Antónia Sabolová-Fabianová, and Anna Jašková. 2020. Explainable classifier for improving the accountability in decision-making for colorectal cancer diagnosis from histopathological images. *Journal of Biomedical Informatics* (2020). doi:10.1016/j.jbi.2020.103523 [CORPUS].

[461] Swati Sachan, Jian-Bo Yang, Dong-Ling Xu, David Eraso Benavides, and Yang Li. 2020. An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications* (2020). doi:10.1016/j.eswa.2019.113100 [CORPUS].

[462] Panda Kumar Sachin and Aaron Schecter. 2024. Advice utilization in combined human-algorithm decision-making: an analysis of preferences and behaviors. *Journal of the Association for Information Systems* (2024). [CORPUS].

[463] Kiarash Sadeghi R., Divesh Ojha, Puneet Kaur, Raj V. Mahto, and Amandeep Dhir. 2024. Explainable artificial intelligence and agile decision-making in supply

chain cyber resilience. *Decision Support Systems* (2024). doi:10.1016/j.dss.2024.114194 [CORPUS].

[464] Sarina Sajjadi Ghaemmaghami and Amirali Salehi-Abari. 2021. DeepGroup: Group Recommendation with Implicit Feedback. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*. doi:10.1145/3459637.3482081 [CORPUS].

[465] Sara Salimzadeh and Ujwal Gadiraju. 2024. When in Doubt! Understanding the Role of Task Characteristics on Peer Decision-Making with AI Assistance. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. doi:10.1145/3627043.3659567 [CORPUS].

[466] Sara Salimzadeh, Gaole He, and Ujwal Gadiraju. 2023. A Missing Piece in the Puzzle: Considering the Role of Task Complexity in Human-AI Decision Making. In *Proceedings of the ACM Conference on User Modeling, Adaptation and Personalization (UMAP)*. doi:10.1145/3565472.3592959 .

[467] Sarah Sandmann, Sarah Riepenhausen, Lucas Plagwitz, and Julian Varghese. 2024. Systematic analysis of ChatGPT, Google search and Llama 2 for clinical decision support tasks. *Nature Communications* (2024). doi:10.1038/s41467-024-46411-8 [CORPUS].

[468] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Ratn Shah. 2021. Quantitative Day Trading from Natural Language using Reinforcement Learning. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*. doi:10.18653/v1/2021.naacl-main.316

[469] Devansh Saxena, Erina Seh-Young Moon, Aryan Chaurasia, Yixin Guan, and Shion Guha. 2023. Rethinking ""Risk"" in Algorithmic Systems Through A Computational Narrative Analysis of Casenotes in Child-Welfare. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581308 [CORPUS].

[470] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias H¨ollerer. 2019. I can do better than your AI: expertise and explanations. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3301275.3302308 [CORPUS].

[471] Nicolas Scharowski, Michaela Benk, Swen J. K¨uhne, Léane Wettstein, and Florian Br¨uhlmann. 2023. Certification Labels for Trustworthy AI: Insights From an Empirical Mixed-Method Study. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3593013.3593994 [CORPUS].

[472] Max Schemmer, Andrea Bartos, Philipp Spitzer, Patrick Hemmer, Niklas K¨uhl, Jonas Liebschner, and Gerhard Satzger. 2023. Towards Effective Human-AI Decision-Making: The Role of Human Learning in Appropriate Reliance on AI Advice. In *International Conference on Information Systems*. [CORPUS].

[473] Max Schemmer, Patrick Hemmer, Maximilian Nitsche, Niklas Kühl, and Michael Vössing. 2022. A Meta-Analysis of the Utility of Explainable Artificial Intelligence in Human-AI Decision-Making. In *AAAI/ACM Conference on AI, Ethics, and Society*. doi:10.1145/3514094.3534128 .

[474] Max Schemmer, Niklas Kuehl, Carina Benz, Andrea Bartos, and Gerhard Satzger. 2023. Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations. In *Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3581641.3584066 [CORPUS].

[475] S¨oren Schleibaum, Lu Feng, Sarit Kraus, and J¨org P. M¨uller. 2024. Adesse: advice explanations in complex repeated decision-making environments. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2024/875 [CORPUS].

[476] Nadine Schlicker, Markus Langer, Sonja K. Ötting, Kevin Baum, Cornelius J. König, and Dieter Wallach. 2021. What to expect from opening up 'black boxes'? Comparing perceptions of justice between human and automated agents. *Computers in Human Behavior* (2021). doi:10.1016/j.chb.2021.106837 [CORPUS].

[477] Timothée Schmude, Laura Koesten, Torsten Möller, and Sebastian Tschiatschek. 2025. Information that matters: exploring information needs of people affected by algorithmic decisions. *International Journal of Human-Computer Studies* (2025). doi:10.1016/j.ijhcs.2024.103380 [CORPUS].

[478] Johannes Schneider and Joshua Peter Handali. 2019. Personalized explanation for machine learning: a conceptualization. In *International Conference on Information Systems (ICIS)*. [CORPUS].

[479] Jakob Schoeffer, Niklas Kuehl, and Yvette Machowski. 2022. "There Is Not Enough Information": On the Effects of Explanations on Perceptions of Informational Fairness and Trustworthiness in Automated Decision-Making. 1616–1628. doi:10.1145/3531146.3533218 [CORPUS].

[480] Tjeerd A.J. Schoonderwoerd, Wiard Jorritsma, Mark A. Neerincx, and Karel van den Bosch. 2021. Human-centered XAI: Developing Design Patterns for Explanations of Clinical Decision Support Systems. *International Journal of Human-Computer Studies* (2021). doi:10.1016/j.ijhcs.2021.102684 [CORPUS].

[481] Philipp Schr¨oppel and Maximilian F¨orster. 2024. Exploring XAI Users' Needs: A Novel Approach to Personalize Explanations Using Contextual Bandits. In *European Conference on Information Systems*. [CORPUS].

[482] Peter Schulam and Suchi Saria. 2017. Reliable decision support using counterfactual models. In *Advances in Neural Information Processing Systems*. [CORPUS].

[483] Candice Schumann, Zhi Lang, Jeffrey Foster, and John Dickerson. 2019. Making the cut: a bandit-based approach to tiered interviewing. In *Advances in Neural Information Processing Systems*. [CORPUS].

[484] Falk Schwendicke, Sarah Mertens, Anselmo Garcia Cantu, Akhilanand Chaurasia, Hendrik Meyer-Lueckel, and Joachim Krois. 2022. Cost-effectiveness of AI for caries detection: randomized trial. *Journal of Dentistry* (2022). doi:10.1016/j.jdent.2022.104080 [CORPUS].

[485] Friso Selten, Marcel Robeer, and Stephan Grimmelikhuijsen. 2023. 'Just like I thought': Street-Level Bureaucrats Trust AI Recommendations if They Confirm Their Professional Judgment. *Public Administration Review* (2023). doi:10.1111/puar.13602 [CORPUS].

[486] Mark Sendak, Madeleine Clare Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. 2020. "The human body is a black box": supporting clinical decision-making with deep learning. In *ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3351095.3372827 [CORPUS].

[487] JooYoung Seo, Sanchita S. Kamath, Aziz Zeidieh, Saairam Venkatesh, and Sean McCurry. 2024. MAIDR Meets AI: Exploring Multimodal LLM-Based Data Visualization Interpretation by and with Blind and Low-Vision Users. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*. doi:10.1145/3663548.3675660 [CORPUS].

[488] Saleh Seyedzadeh, Farzad Pour Rahimian, Stephen Oliver, Sergio Rodriguez, and Ivan Glesk. 2020. Machine learning modelling for predicting non-domestic buildings energy performance: a model to support deep energy retrofit decision-making. *Applied Energy* (2020). doi:10.1016/j.apenergy.2020.115908 [CORPUS].

[489] Ankit Shah, Rajesh Ganesan, Sushil Jajodia, Pierangela Samarati, and Hasan Cam. 2020. Adaptive Alert Management for Balancing Optimal Performance among Distributed CSOCs using Reinforcement Learning. *IEEE Transactions on Parallel and Distributed Systems* (2020). doi:10.1109/TPDS.2019.2927977 [CORPUS].

[490] Md Shajalal, Alexander Boden, and Gunnar Stevens. 2022. Explainable product backorder prediction exploiting CNN: Introducing explainable models in businesses. *Electronic Markets* (2022). doi:10.1007/s12525-022-00599-z [CORPUS].

[491] Ameneh Shamekhi, Q. Vera Liao, Dakuo Wang, Rachel K. E. Bellamy, and Thomas Erickson. 2018. Face Value? Exploring the Effects of Embodiment for a Group Facilitation Agent. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3173574.3173965

[492] Siqing Shan and Yinong Li. 2024. Research on the application framework of generative AI in emergency response decision support systems for emergencies. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2423335 [CORPUS].

[493] Ruoxi Shang, K. J. Kevin Feng, and Chirag Shah. 2022. Why Am I Not Seeing It? Understanding Users' Needs for Counterfactual Explanations in Everyday Recommendations. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3531146.3533189 [CORPUS].

[494] Daniel B. Shank, Alyssa DeSanti, and Timothy Maninger. 2019. When are artificial intelligence versus human agents faulted for wrongdoing? Moral attributions after individual and joint decisions. *Information, Communication & Society* (2019). doi:10.1080/1369118X.2019.1568515 [CORPUS].

[495] James C. Shanteau. 1970. An additive model for sequential decision making. *Journal of Experimental Psychology* 85, 2 (1970), 181–191. doi:10.1037/h0029552

[496] Shavneet Sharma, Nazrul Islam, Gurmeet Singh, and Amandeep Dhir. 2024. Why Do Retail Customers Adopt Artificial Intelligence (AI) Based Autonomous Decision-Making Systems? *IEEE Transactions on Engineering Management* (2024). doi:10.1109/TEM.2022.3157976 [CORPUS].

[497] Jiang Shen, Fusheng Liu, Man Xu, Lipeng Fu, Zhenhe Dong, and Jiachao Wu. 2022. Decision support analysis for risk identification and control of patients affected by COVID-19 based on Bayesian networks. *Expert Systems with Applications* (2022). doi:10.1016/j.eswa.2022.116547 [CORPUS].

[498] Lin Sheng, Zhenyu Gu, and Fangyuan Chang. 2024. A novel integration strategy for uncertain knowledge in group decision-making with artificial opinions: a dsfit-soa-dematel approach. *Expert Systems with Applications* (2024). doi:10.1016/j.eswa.2023.122886 [CORPUS].

[499] Si Shi, Yuhuang Gong, and Dogan Gursoy. 2021. Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic–systematic model. *Journal of Travel Research* (2021). doi:10.1177/0047287520966395 [CORPUS].

[500] Si Shi, Jianjun Li, Guohui Li, Peng Pan, Qi Chen, and Qing Sun. 2022. GPM: A Graph Convolutional Network Based Reinforcement Learning Framework for Portfolio Management. *Neurocomputing* (2022). doi:10.1016/j.neucom.2022.04.105 [CORPUS].

[501] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: implications for explainable AI. *International Journal of Human-Computer Studies* (2021). doi:10.1016/j.ijhcs.2020.102551 [CORPUS].

[502] Joo Hwan Shin, Junmo Kwon, Jong Uk Kim, Hyewon Ryu, Jehyung Ok, S. Joon Kwon, Hyunjin Park, and Tae-il Kim. 2022. Wearable EEG electronics for a Brain–AI Closed-Loop System to enhance autonomous machine decision-making. *npj Flexible Electronics* (2022). doi:10.1038/s41528-022-00164-w [CORPUS].

[503] Galit Shmueli and Soumya Ray. 2024. Reimagining the journal editorial process: an AI-augmented versus an AI-driven future. *Journal of the Association for Information Systems* (2024). [CORPUS].

[504] Avital Shulner-Tal, Tsvi Kuflik, and Doron Kliger. 2023. Enhancing fairness perception – towards human-centred AI and personalized explanations understanding the factors influencing laypeople's fairness perceptions of algorithmic decisions. *International Journal of Human–Computer Interaction* (2023). doi:10.1080/10447318.2022.2095705 [CORPUS].

[505] Avital Shulner-Tal, Tsvi Kuflik, Doron Kliger, and Azzurra Mancini. 2024. Who Made That Decision and Why? Users' Perceptions of Human Versus AI Decision-Making and the Power of Explainable-AI. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2348843 [CORPUS].

[506] Chenglei Si, Navita Goyal, Tongshuang Wu, Chen Zhao, Shi Feng, Hal Daumé Iii, and Jordan Boyd-Graber. 2024. Large Language Models Help Humans Verify Truthfulness – Except When They Are Convincingly Wrong. In *Proceedings of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*. doi:10.18653/v1/2024.naacl-long.81 [CORPUS].

[507] Matthew Sidji, Wally Smith, and Melissa J. Rogerson. 2023. The Hidden Rules of Hanabi: How Humans Outperform AI Agents. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581550 [CORPUS].

[508] Chonghao Sima, Katrin Renz, Kashyap Chitta, Li Chen, Hanxue Zhang, Chengen Xie, Jens Beißwenger, Ping Luo, Andreas Geiger, and Hongyang Li. 2025. DriveLM: Driving with Graph Visual Question Answering. In *European Conference on Computer Vision*. doi:10.1007/978-3-031-72943-0_15 [CORPUS].

[509] Prashant Kumar Singh and Prabir Sarkar. 2023. An artificial neural network tool to support the decision making of designers for environmentally conscious product development. *Expert Systems with Applications* (2023). doi:10.1016/j.eswa.2022.118679 [CORPUS].

[510] Siddharth Singi, Zhanpeng He, Alvin Pan, Sandip Patel, Gunnar A. Sigurdsson, Robinson Piramuthu, Shuran Song, and Matei Ciocarlie. 2024. Decision Making for Human-in-the-loop Robotic Agents via Uncertainty-Aware Reinforcement Learning. In *IEEE International Conference on Robotics and Automation*. doi:10.1109/ICRA57147.2024.10611425 [CORPUS].

[511] Sumedha Singla, Motahhare Eslami, Brian Pollack, Stephen Wallace, and Kayhan Batmanghelich. 2023. Explaining the black-box smoothly—a counterfactual approach. *Medical Image Analysis* (2023). doi:10.1016/j.media.2022.102721 [CORPUS].

[512] Venkatesh Sivaraman, Leigh A Bukowski, Joel Levin, Jeremy M. Kahn, and Adam Perer. 2023. Ignore, Trust, or Negotiate: Understanding Clinician Acceptance of AI-Based Treatment Recommendations in Health Care. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581075 [CORPUS].

[513] Andria Smith, Hunter Phoenix van Wagoner, Ksenia Keplinger, and Can Celebi. 2025. Navigating AI convergence in human–artificial intelligence teams: A signaling theory approach. *Journal of Organizational Behavior* (2025). doi:10.1002/job.2856 [CORPUS].

[514] Wonyoung So, Pranay Lohia, Rakesh Pimplikar, A. E. Hosoi, and Catherine D'Ignazio. 2022. Beyond Fairness: Reparative Algorithms to Address Historical Injustices of Housing Discrimination in the US. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3531146.3533160 [CORPUS].

[515] Dalia Sobhy, Leandro Minku, Rami Bahsoon, and Rick Kazman. 2022. Continuous and Proactive Software Architecture Evaluation: An IoT Case. *ACM Transactions on Software Engineering and Methodology* (2022). doi:10.1145/3492762 [CORPUS].

[516] Kacper Sokol and Peter Flach. 2018. Glass-box: explaining AI decisions with counterfactual statements through conversation with a voice-enabled virtual assistant. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2018/865 [CORPUS].

[517] Taylor Sorensen, Liwei Jiang, Jena D. Hwang, Sydney Levine, Valentina Pyatkin, Peter West, Nouha Dziri, Ximing Lu, Kavel Rao, Chandra Bhagavatula, Maarten Sap, John Tasioulas, and Yejin Choi. 2024. Value Kaleidoscope: Engaging AI with Pluralistic Human Values, Rights, and Duties. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v38i18.29970 [CORPUS].

[518] Pierre Stock and Moustapha Cisse. 2018. ConvNets and ImageNet Beyond Accuracy: Understanding Mistakes and Uncovering Biases. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

[519] Eleni Straitouri and Manuel Gomez Rodriguez. 2024. Designing Decision Support Systems Using Counterfactual Prediction Sets. In *International Conference on Machine Learning (ICML)*. [CORPUS].

[520] Katherine J. Strandburg. 2019. Rulemaking and inscrutable automated decision tools. *Columbia Law Review* (2019). doi:10.2307/26810852 [CORPUS].

[521] Franz Strich, Anne-Sophie Mayer, and Marina Fiedler. 2021. What do I do in a world of artificial intelligence? Investigating the impact of substitutive decision-making AI systems on employees' professional role identity. *Journal of the Association for Information Systems* (2021). [CORPUS].

[522] Jiao Sun, Q. Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D. Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3490099.3511119 [CORPUS].

[523] Lu Sun, Ziqian Liu, Zhaolong Ning, Jie Wang, and Xianping Fu. 2024. Multi-Agent Q-Net Enhanced Coevolutionary Algorithm for Resource Allocation in Emergency Human-Machine Fusion UAV-MEC System. *IEEE Transactions on Automation Science and Engineering* (2024). doi:10.1109/TASE.2024.3409551 [CORPUS].

[524] Lu Sun, Stone Tao, Junjie Hu, and Steven P. Dow. 2024. MetaWriter: Exploring the Potential and Perils of AI Writing Support in Scientific Peer Review. *Proceedings of the ACM on Human-Computer Interaction* (2024). doi:10.1145/3637371 [CORPUS].

[525] Viktor Suter, Miriam Meckel, Morteza Shahrezaye, and Léa Steinacker. 2022. AI Suffrage: A four-country survey on the acceptance of an automated voting system. (2022). [CORPUS].

[526] Siddharth Swaroop, Zana Buçinca, Krzysztof Z. Gajos, and Finale Doshi-Velez. 2024. Accuracy-Time Tradeoffs in AI-Assisted Decision Making under Time Pressure. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3640543.3645206 [CORPUS].

[527] Ran Tan, Yang Li, and Qian Huang. 2021. Enhancing service chatbot effectiveness: the effect of dyadic communication traits on consumer unplanned purchase. In *International Conference on Information Systems*. [CORPUS].

[528] Xiaolin Tang, Guichuan Zhong, Shen Li, Kai Yang, Keqi Shu, Dongpu Cao, and Xianke Lin. 2023. Uncertainty-Aware Decision-Making for Autonomous Driving at Uncontrolled Intersections. *IEEE Transactions on Intelligent Transportation Systems* (2023). doi:10.1109/TITS.2023.3283019 [CORPUS].

[529] Zhenwei Tang, Difan Jiao, Reid McIlroy-Young, Jon Kleinberg, Siddhartha Sen, and Ashton Anderson. 2024. Maia-2: a unified model for human-ai alignment in chess. In *Advances in Neural Information Processing Systems*. [CORPUS].

[530] Anna Taudien, Andreas F¨ugener, Alok Gupta, and Wolfgang Ketter. 2022. The effect of AI advice on human confidence in decision-making. In *Hawaii International Conference on System Sciences*. [CORPUS].

[531] Heliodoro Tejeda, Aakriti Kumar, Padhraic Smyth, and Mark Steyvers. 2022. AI-Assisted Decision-making: a Cognitive Modeling Approach to Infer Latent Reliance Strategies. *Computational Brain & Behavior* (2022). doi:10.1007/s42113-022-00157-y [CORPUS].

[532] Angela Testi and Elena Tànfani. 2009. Tactical and operational decisions for operating room planning: Efficiency and welfare implications. *Health Care Management Science* 12 (2009). doi:10.1007/s10729-008-9093-4

[533] Navamayooran Thavanesan, Ganesh Vigneswaran, Indu Bodala, and Timothy J. Underwood. 2023. The Oesophageal Cancer Multidisciplinary Team: Can Machine Learning Assist Decision-Making? *Journal of Gastrointestinal Surgery* (2023). doi:10.1007/s11605-022-05575-8 .

[534] P. C. Trimmer, A. I. Houston, J. A. R. Marshall, Rafal Bogacz, E. S. Paul, M. T. Mendl, and J. M. McNamara. 2011. Decision-making under uncertainty: biases and Bayesians. *Animal Cognition* 14, 4 (2011), 465–476. doi:10.1007/s10071-011-0387-4

[535] Wan-Lun Tsai, Li-Wen Su, Tsai-Yen Ko, Tse-Yu Pan, and Min-Chun Hu. 2021. Feasibility Study on Using AI and VR for Decision-Making Training of Basketball Players. *IEEE Transactions on Learning Technologies* (2021). doi:10.1109/TLT.2022.3145093 [CORPUS].

[536] Philipp Tschandl, Christoph Rinner, Zoe Apalla, Giuseppe Argenziano, Noel Codella, Allan Halpern, Monika Janda, Aimilios Lallas, Caterina Longo, Josep Malvehy, John Paoli, Susana Puig, Cliff Rosendahl, H. Peter Soyer, Iris Zalaudek, and Harald Kittler. 2020. Human–computer collaboration for skin cancer recognition. *Nature Medicine* (2020). doi:10.1038/s41591-020-0942-0 [CORPUS].

[537] Naoum Tsolakis, Dimitris Zissis, Spiros Papaefthimiou, and Nikolaos Korfiatis. 2022. Towards AI Driven Environmental Sustainability: An Application of Automated Logistics in Container Port Terminals. *International Journal of Production Research* 60 (2022).

[538] Abdullah Aman Tutul, Ehsanul Haque Nirjhar, and Theodora Chaspari. 2024. Investigating trust in human-ai collaboration for a speech-based data analytics task. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2328910 [CORPUS].

[539] Jesper Tveit, Harald Aurlien, Sergey Plis, Vince D. Calhoun, William O. Tatum, Donald L. Schomer, Vibeke Arntsen, Fieke Cox, Firas Fahoum, William B. Gallentine, Elena Gardella, Cecil D. Hahn, Aatif M. Husain, Sudha Kessler, Mustafa Aykut Kural, Fábio A. Nascimento, Hatice Tankisi, Line B. Ulvin, Richard Wennberg, and Sándor Beniczky. 2023. Automated interpretation of clinical electroencephalograms using artificial intelligence. *Journal of the American Medical Association Neurology* (2023). doi:10.1001/jamaneurol.2023.1645 [CORPUS].

[540] Nichole S. Tyler, Clara M. Mosquera-Lopez, Leah M. Wilson, Robert H. Dodier, Deborah L. Branigan, Virginia B. Gabo, Florian H. Guillot, Wade W. Hilts, Joseph El Youssef, Jessica R. Castle, and Peter G. Jacobs. 2020. An artificial intelligence decision support system for the management of type 1 diabetes. *Nature Metabolism* (2020). doi:10.1038/s42255-020-0212-y [CORPUS].

[541] Suleyman Uslu, Davinder Kaur, Samuel J. Rivera, Arjan Durresi, Meghna Babbar-Sebens, and Jenna H. Tilt. 2024. A Trustworthy and Responsible Decision-Making Framework for Resource Management in Food-Energy-Water Nexus: A Control-Theoretical Approach. *ACM Transactions on Intelligent Systems and Technology* (2024). doi:10.1145/3660640 [CORPUS].

[542] Michelle Vaccaro, Abdullah Almaatouq, and Thomas Malone. 2024. When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour* (2024). doi:10.1038/s41562-024-02024-1 .

[543] Ehsan Vakili, Abdollah Amirkhani, and Behrooz Mashadi. 2024. Dqn-based ethical decision-making for self-driving cars in unavoidable crashes: an applied ethical knob. *Expert Systems with Applications* (2024). doi:10.1016/j.eswa.2024.124569 [CORPUS].

[544] Gilmer Valdes, Charles B. Simone, Josephine Chen, Alexander Lin, Sue S. Yom, Adam J. Pattison, Colin M. Carpenter, and Timothy D. Solberg. 2017. Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making. *Radiotherapy and Oncology* (2017). doi:10.1016/j.radonc.2017.10.014 [CORPUS].

[545] Karthik Valmeekam, Sarath Sreedharan, Sailik Sengupta, and Subbarao Kambhampati. 2021. RADAR-X: An Interactive Interface Pairing Contrastive Explanations with Revised Plan Suggestions. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v35i18.18009 [CORPUS].

[546] Colin van Noordt and Gianluca Misuraca. 2022. Artificial Intelligence for the Public Sector: Results of Landscaping the Use of AI in Government across the European Union. *Government Information Quarterly* (2022). doi:10.1016/j.giq.2022.101714 [CORPUS].

[547] Helena Vasconcelos, Matthew J¨orke, Madeleine Grunde-McLaughlin, Tobias Gerstenberg, Michael S. Bernstein, and Ranjay Krishna. 2023. Explanations Can Reduce Overreliance on AI Systems During Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 129 (April 2023), 38 pages. doi:10.1145/3579605 [CORPUS].

[548] Baptiste Vasey, Myura Nagendran, Bruce Campbell, David A. Clifton, Gary S. Collins, Spiros Denaxas, Alastair K. Denniston, Livia Faes, Bart Geerts, Mudathir Ibrahim, Xiaoxuan Liu, Bilal A. Mateen, Piyush Mathur, Melissa D. McCradden, Lauren Morgan, Johan Ordish, Campbell Rogers, Suchi Saria, Daniel S. W. Ting, Peter Watkinson, Wim Weber, Peter Wheatstone, Peter McCulloch, and the DECIDE-AI expert group. 2022. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine* (2022). doi:10.1038/s41591-022-01772-9 [CORPUS].

[549] Baptiste Vasey, Stephan Ursprung, Benjamin Beddoe, Elliott H. Taylor, Neale Marlow, Nicole Bilbro, Peter Watkinson, and Peter McCulloch. 2021. Association of clinician diagnostic performance with machine learning–based decision support systems: a systematic review. *JAMA Network Open* 4, 3 (2021), e211276. doi:10.1001/jamanetworkopen.2021.1276 .

[550] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3173574.3174014 [CORPUS].

[551] Erik Veitch, Henrikke Dybvik, Martin Steinert, and Ole Andreas Alsos. 2024. Collaborative Work with Highly Automated Marine Navigation Systems. *Computer Supported Cooperative Work (CSCW)* (2024). doi:10.1007/s10606-022-09450-7 [CORPUS].

[552] Tobias Vente. 2023. Advancing Automation of Design Decisions in Recommender System Pipelines. In *ACM Conference on Recommender Systems (RecSys)*. doi:10.1145/3604915.3608886 [CORPUS].

[553] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021. How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction* (2021). doi:10.1145/3476068 .

[554] Maria Virvou, George A. Tsihrintzis, and Evangelia-Aikaterini Tsichrintzi. 2024. Virtsi: a novel trust dynamics model enhancing artificial intelligence collaboration with human users – insights from a chatgpt evaluation study. *Information Sciences* (2024). doi:10.1016/j.ins.2024.120759 [CORPUS].

[555] Charles AJ Vlek. 1996. A multi-level, multi-stage and multi-attribute perspective on risk assessment, decision-making and risk control. *Risk decision and Policy* 1, 1 (1996), 9–31.

[556] Paul Walsh, Justin Thornton, Julie Asato, Nicholas Walker, Gary McCoy, Joe Baal, Jed Baal, Nanse Mendoza, and Faried Banimahd. 2014. Approaches to describing inter-rater reliability of the overall clinical appearance of febrile infants and toddlers in the emergency department. *PeerJ* 2 (2014), e651. doi:10.7717/peerj.651

[557] Chengbo Wang, Xinyu Zhang, Hongbo Gao, Musa Bashir, Huanhuan Li, and Zaili Yang. 2024. Colergs-constrained safe reinforcement learning for realising mass's risk-informed collision avoidance decision making. *Knowledge-Based Systems* (2024). doi:10.1016/j.knosys.2024.112205 [CORPUS].

[558] Clinton J. Wang, Charlie A. Hamm, Lynn J. Savic, Marc Ferrante, Isabel Schobert, Todd Schlachter, MingDe Lin, Jeffrey C. Weinreb, James S. Duncan, Julius Chapiro, and Brian Letzen. 2019. Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features. *European Radiology* (2019). doi:10.1007/s00330-019-06214-8 [CORPUS].

[559] Dakuo Wang, Liuping Wang, Zhan Zhang, Ding Wang, Haiyi Zhu, Yvonne Gao, Xiangmin Fan, and Feng Tian. 2021. "Brilliant AI Doctor" in Rural Clinics: Challenges in AI-Powered Clinical Decision Support System Deployment. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445432 [CORPUS].

[560] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Conference on Human Factors in Computing Systems*. doi:10.1145/3290605.3300831 [CORPUS].

[561] Ge Wang, Yue Guo, Weimin Zhang, Shenghua Xie, and Qiwei Chen. 2023. What type of algorithm is perceived as fairer and more acceptable? A comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs. *Government Information Quarterly* (2023). doi:10.1016/j.giq.2023.101803 [CORPUS].

[562] Ge Wang, Shenghua Xie, and Xiaoqian Li. 2024. Artificial intelligence, types of decisions, and street-level bureaucrats: evidence from a survey experiment. *Public Management Review* (2024). doi:10.1080/14719037.2022.2070243 [CORPUS].

[563] Junjie Wang, Qichao Zhang, Dongbin Zhao, and Yaran Chen. 2019. Lane Change Decision-making through Deep Reinforcement Learning with Rule-based Constraints. In *International Joint Conference on Neural Networks*. doi:10.1109/IJCNN.2019.8852110 [CORPUS].

[564] Ping Wang and Heng Ding. 2024. The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-ai trust and decision performance. *Information Processing & Management* (2024). doi:10.1016/j.ipm.2024.103732 [CORPUS].

[565] Wei Wang, Christopher Lesner, Alexander Ran, Marko Rukonic, Jason Xue, and Eric Shiu. 2020. Using Small Business Banking Data for Explainable Credit Risk Scoring. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v34i08.7055 [CORPUS].

[566] Wei-Yao Wang, Teng-Fong Chan, Wen-Chih Peng, Hui-Kuo Yang, Chih-Chuan Wang, and Yao-Chung Fan. 2022. How Is the Stroke? Inferring Shot Influence in Badminton Matches via Long Short-term Dependencies. *ACM Transactions on Intelligent Systems and Technology* (2022). doi:10.1145/3551391 [CORPUS].

[567] Xinru Wang, Chen Liang, and Ming Yin. 2023. The Effects of AI Biases and Explanations on Human Decision Fairness: A Case Study of Bidding in Rental Housing Markets. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2023/343 [CORPUS].

[568] Xinru Wang and Ming Yin. 2021. Are Explanations Helpful? A Comparative Study of the Effects of Explanations in AI-Assisted Decision-Making. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3397481.3450650 [CORPUS].

[569] Xinru Wang and Ming Yin. 2023. Watch Out for Updates: Understanding the Effects of Model Explanation Updates in AI-Assisted Decision Making. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581366 [CORPUS].

[570] Yixin Wang, Dhanya Sridhar, and David Blei. 2024. Adjusting machine learning decisions for equal opportunity and counterfactual fairness. *Transactions on Machine Learning Research* (2024). [CORPUS].

[571] Jonas Wanner, Lukas-Valentin Herm, Kai Heinrich, and Christian Janiesch. 2022. The effect of transparency and trust on intelligent system acceptance: Evidence from a user-based study. *Electronic Markets* (2022). doi:10.1007/s12525-022-00593-5 [CORPUS].

[572] Paul Ward. 2023. Choice, Uncertainty, and Decision Superiority: Is Less AI-Enabled Decision Support More? *IEEE Transactions on Human-Machine Systems* (2023). doi:10.1109/THMS.2023.3279036 .

[573] Greta Warren, Ruth M.J. Byrne, and Mark T. Keane. 2024. Categorical and Continuous Features in Counterfactual Explanations of AI Systems. *ACM Transactions on Interactive Intelligent Systems* (2024). doi:10.1145/3673907 [CORPUS].

[574] Sebastian Weber, Hans Christian Klein, Dominik Siemon, Bastian Kordyaka, and Bj¨orn Niehaves. 2024. Designing successful human-ai collaboration for creative-problem solving in architectural design. In *International Conference on Information Systems*. [CORPUS].

[575] Nilmini Pradeepika Weerasinghe, Rebecca Jing Yang, and Chen Wang. 2022. Learning from success: a machine learning approach to guiding solar building envelope applications in non-domestic market. *Journal of Cleaner Production* (2022). doi:10.1016/j.jclepro.2022.133997 [CORPUS].

[576] Alexander Wei, Nika Haghtalab, and Jacob Steinhardt. 2023. Jailbroken: How Does LLM Safety Training Fail?. In *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. 80079–80110.

[577] Lijun Wei, Derek R. Magee, Vania Dimitrova, Barry Clarke, Heshan Du, Quratul-ain Mahesar, Kareem Al Ammari, and Anthony G. Cohn. 2018. Automated reasoning for city infrastructure maintenance decision support. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2018/868 [CORPUS].

[578] Bin Weng, Mohamed A. Ahmed, and Fadel M. Megahed. 2017. Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications* (2017). doi:10.1016/j.eswa.2017.02.041 [CORPUS].

[579] Monika Westphal, Michael V¨ossing, Gerhard Satzger, Galit B. Yom-Tov, and Anat Rafaeli. 2023. Decision control and explanations in human-ai collaboration: improving user perceptions and compliance. *Computers in Human Behavior* (2023). doi:10.1016/j.chb.2023.107714 [CORPUS].

[580] Ann M. Wieben, Bader G. Alreshidi, Brian J. Douthit, Marisa Sileo, Pankaj Vyas, Linsey Steege, and Andrea Gilmore-Bykovskyi. 2024. Nurses' perceptions of the design, implementation, and adoption of machine learning clinical decision support: a descriptive qualitative study. *Journal of Nursing Scholarship* (2024). doi:10.1111/jnu.13001 [CORPUS].

[581] Carolin Wienrich and Marc Erich Latoschik. 2021. eXtended Artificial Intelligence: New Prospects of Human-AI Interaction Research. *Frontiers in Virtual Reality* 2 (2021), 686783. doi:10.3389/frvir.2021.686783

[582] Christopher Y. K. Williams, Brenda Y. Miao, Aaron E. Kornblith, and Atul J. Butte. 2024. Evaluating the use of large language models to provide clinical recommendations in the Emergency Department. *Nature Communications* (2024). doi:10.1038/s41467-024-52415-1 [CORPUS].

[583] Magdalena Wischnewski, Nicole Krämer, and Emmanuel Müller. 2023. Measuring and Understanding Trust Calibrations for Automated Systems: A Survey of the State-Of-The-Art and Future Directions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, Article 755. doi:10.1145/3544548.3581197

[584] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* (2016). doi:10.1145/2907069

[585] Kyle Hollins Wray, Luis Pineda, and Shlomo Zilberstein. 2016. Hierarchical approach to transfer of control in semi-autonomous systems. In *Proceedings of the International Joint Conference on Artificial Intelligence*. [CORPUS].

[586] Jingda Wu, Ziyou Song, and Chen Lv. 2024. Deep Reinforcement Learning-Based Energy-Efficient Decision-Making for Autonomous Electric Vehicle in Dynamic Traffic Environments. *IEEE Transactions on Transportation Electrification* (2024). doi:10.1109/TTE.2023.3290069 [CORPUS].

[587] Oskar Wysocki, Jessica Katharine Davies, Markel Vigo, Anne Caroline Armstrong, Dónal Landers, Rebecca Lee, and André Freitas. 2023. Assessing the communication gap between AI models and healthcare professionals: explainability, utility and trust in AI-driven clinical decision-making. *Artificial Intelligence* (2023). doi:10.1016/j.artint.2022.103839 [CORPUS].

[588] Yu Xiong, Runze Wu, Shiwei Zhao, Jianrong Tao, Xudong Shen, Tangjie Lyu, Changjie Fan, and Peng Cui. 2023. A Data-Driven Decision Support Framework for Player Churn Analysis in Online Games. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3580305.3599759 [CORPUS].

[589] Cheng Xu, Hao Xu, Yanqi Sun, and Wanfang Xiong. 2024. The digital siren's call: accepting unethical AI advice. *International Journal of Human–Computer Interaction* (2024). doi:10.1080/10447318.2024.2400396 [CORPUS].

[590] Xin Xu, Lei Zuo, Xin Li, Lilin Qian, Junkai Ren, and Zhenping Sun. 2020. A Reinforcement Learning Approach to Autonomous Decision Making of Intelligent Vehicles on Highways. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50, 8 (2020), 2920–2934. doi:10.1109/TSMC.2018.2870983 [CORPUS].

[591] Zhun Xu, Liyun Xu, Xufeng Ling, and Beikun Zhang. 2023. Data-driven hierarchical learning and real-time decision-making of equipment scheduling and location assignment in automatic high-density storage systems. *International Journal of Production Research* (2023). doi:10.1080/00207543.2022.2148011 [CORPUS].

[592] Zelai Xu, Chao Yu, Fei Fang, Yu Wang, and Yi Wu. 2024. Language agents with reinforcement learning for strategic play in the werewolf game. In *International Conference on Machine Learning (ICML)*. [CORPUS].

[593] Ziqi Xu, Jingwen Zhang, Jacob Greenberg, Madelyn Frumkin, Saad Javeed, Justin K. Zhang, Braeden Benedict, Kathleen Botterbush, Thomas L. Rodebaugh, Wilson Z. Ray, and Chenyang Lu. 2024. Predicting Multi-dimensional Surgical Outcomes with Multi-modal Mobile Sensing: A Case Study with Patients Undergoing Lumbar Spine Surgery. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2024). doi:10.1145/3659628 [CORPUS].

[594] Fumeng Yang, Zhuanyi Huang, Jean Scholtz, and Dustin L. Arendt. 2020. How do visual explanations foster end users' appropriate trust in machine learning?. In *International Conference on Intelligent User Interfaces (IUI)*. doi:10.1145/3377325.3377480 [CORPUS].

[595] Haoyu Yang, Kaiming Xiao, Lihua Liu, Hongbin Huang, and Weiming Zhang. 2022. An online learning approach towards far-sighted emergency relief planning under intentional attacks in conflict areas. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2022/649 [CORPUS].

[596] Kai Yang, Xiaolin Tang, Sen Qiu, Shufeng Jin, Zichun Wei, and Hong Wang. 2023. Towards Robust Decision-Making for Autonomous Driving on Highway. *IEEE Transactions on Vehicular Technology* (2023). doi:10.1109/TVT.2023.3268500 [CORPUS].

[597] Qian Yang, Yuexing Hao, Kexin Quan, Stephen Yang, Yiran Zhao, Volodymyr Kuleshov, and Fei Wang. 2023. Harnessing Biomedical Literature to Calibrate Clinicians' Trust in AI Decision Support Systems. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581393 [CORPUS].

[598] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (2020), 1–13. doi:10.1145/3313831.3376301

[599] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Conference on Human Factors in Computing Systems*. doi:10.1145/3290605.3300468 [CORPUS].

[600] Scott Cheng-Hsin Yang, Nils Erik Tomas Folke, and Patrick Shafto. 2022. A psychological theory of explainability. In *International Conference on Machine Learning (ICML)*. [CORPUS].

[601] Xiaoxi Yao, David R. Rushlow, Jonathan W. Inselman, Rozalina G. McCoy, Thomas D. Thacher, Emma M. Behnken, Matthew E. Bernard, Steven L. Rosas, Abdulla Akfaly, Artika Misra, Paul E. Molling, Joseph S. Krien, Randy M. Foss, Barbara A. Barry, Konstantinos C. Siontis, Suraj Kapa, Patricia A. Pellikka, Francisco Lopez-Jimenez, Zachi I. Attia, Nilay D. Shah, Paul A. Friedman, and Peter A. Noseworthy. 2021. Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nature Medicine* (2021). doi:10.1038/s41591-021-01335-4 [CORPUS].

[602] Randy Yeh, Jennifer H. Kuo, Bernice Huang, Parnian Shobeiri, James A. Lee, Yu-Kwang Donovan Tay, Gaia Tabacco, John P. Bilezikian, and Laurent Dercle. 2024. Machine learning-derived clinical decision algorithm for the diagnosis of hyperfunctioning parathyroid glands in patients with primary hyperparathyroidism. *European Radiology* (2024). doi:10.1007/s00330-024-11159-8 [CORPUS].

[603] Hongxu Yin and Niraj K. Jha. 2017. A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles. *IEEE Transactions on Multi-Scale Computing Systems* (2017). doi:10.1109/TMSCS.2017.2710194 [CORPUS].

[604] Ryosuke Yokoi and Kazuya Nakayachi. 2021. Trust in autonomous cars: exploring the role of shared moral values, reasoning, and emotion in safety-critical decisions. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 63, 6 (2021), 1004–1019. doi:10.1177/0018720820933041 [CORPUS].

[605] Tae Keun Yoo, Ik Hee Ryu, Geunyoung Lee, Youngnam Kim, Jin Kuk Kim, In Sik Lee, Jung Sub Kim, and Tyler Hyungtaek Rim. 2019. Adopting machine learning to automatically identify candidate patients for corneal refractive surgery. *Nature Partner Journals Digital Medicine* (2019). doi:10.1038/s41746-019-0135-8 [CORPUS].

[606] Jinsung Yoon, Ahmed Alaa, Scott Hu, and Mihaela Schaar. 2016. Forecasticu: a prognostic decision support system for timely prediction of intensive care unit admission. In *International Conference on Machine Learning*. [CORPUS].

[607] Yue You, Yubo Kou, Xianghua Ding, and Xinning Gui. 2021. The Medical Authority of AI: A Study of AI-enabled Consumer-Facing Health Technology. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3411764.3445657 [CORPUS].

[608] Hongbo Yu, Jenifer Z. Siegel, and Molly J. Crockett. 2018. Modeling morality in 3-d: decision-making, judgment, and inference. *Topics in Cognitive Science* (2018). doi:10.1111/tops.12382 [CORPUS].

[609] You Yu, Jiahong Li, Samuel A. Solomon, Jihong Min, Jiaobing Tu, Wei Guo, Changhao Xu, Yu Song, and Wei Gao. 2022. All-printed soft human-machine interface for robotic physicochemical sensing. *Science Robotics* (2022). [CORPUS].

[610] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P. Gummadi. 2019. Fairness constraints: a flexible approach for fair classification. *Journal of Machine Learning Research* (2019). [CORPUS].

[611] Muhammad Bilal Zafar, Isabel Valera, Manuel Rodriguez, Krishna Gummadi, and Adrian Weller. 2017. From parity to preference-based notions of fairness in classification. In *Advances in Neural Information Processing Systems*. [CORPUS].

[612] Guanglu Zhang, Leah Chong, Kenneth Kotovsky, and Jonathan Cagan. 2023. Trust in an AI versus a Human Teammate: The Effects of Teammate Identity and Performance on Human-AI Cooperation. *Computers in Human Behavior* (2023). doi:10.1016/j.chb.2022.107536 [CORPUS].

[613] Junzhe Zhang and Elias Bareinboim. 2024. Fairness in Decision-Making — The Causal Explanation Formula. In *Proceedings of the AAAI Conference on Artificial Intelligence*. doi:10.1609/aaai.v32i1.11564 [CORPUS].

[614] Junjie Zhang, Yupeng Hou, Ruobing Xie, Wenqi Sun, Julian McAuley, Wayne Xin Zhao, Leyu Lin, and Ji-Rong Wen. 2024. AgentCF: Collaborative Learning with Autonomous Language Agents for Recommender Systems. In *The ACM Web Conference*. doi:10.1145/3589334.3645537 [CORPUS].

[615] Lingyu Zhang, Zhengran Ji, Nicholas R Waytowich, and Boyuan Chen. 2024. Guide: real-time human-shaped agents. In *Advances in Neural Information Processing Systems*. [CORPUS].

[616] Qianqian Zhang, Yu Kang, Yun-Bo Zhao, Pengfei Li, and Shiyi You. 2022. Traded Control of Human–Machine Systems for Sequential Decision-Making Based on Reinforcement Learning. *IEEE Transactions on Artificial Intelligence* (2022). doi:10.1109/TAI.2021.3127857 [CORPUS].

[617] Qiaoning Zhang, Matthew L Lee, and Scott Carter. 2022. You Complete Me: Human-AI Teams and Complementary Expertise. In *ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3491102.3517791 [CORPUS].

[618] Qi Zhang, Chao Xu, Jie Li, Yicheng Sun, Jinsong Bao, and Dan Zhang. 2024. Llm-tsfd: an industrial time series human-in-the-loop fault diagnosis method based on a large language model. *Expert Systems with Applications* (2024). doi:10.1016/j.eswa.2024.125861 [CORPUS].

[619] Rui Zhang, Christopher Flathmann, Geoff Musick, Beau Schelble, Nathan J. McNeese, Bart Knijnenburg, and Wen Duan. 2024. I Know This Looks Bad, But I Can Explain: Understanding When AI Should Explain Actions In Human-AI Teams. *ACM Transactions on Interactive Intelligent Systems* (2024). doi:10.1145/3635474 [CORPUS].

[620] Raina Zexuan Zhang, Ellie J. Kyung, Chiara Longoni, Luca Cian, and Kellen Mrkva. 2025. Ai-induced indifference: unfair AI reduces prosociality. *Cognition* (2025). doi:10.1016/j.cognition.2024.105937 [CORPUS].

[621] Shao Zhang, Jianing Yu, Xuhai Xu, Changchang Yin, Yuxuan Lu, Bingsheng Yao, Melanie Tory, Lace M. Padilla, Jeffrey Caterino, Ping Zhang, and Dakuo Wang. 2024. Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis. In *CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3613904.3642343 [CORPUS].

[622] Wenbin Zhang and Eirini Ntoutsi. 2019. Faht: an adaptive fairness-aware decision tree classifier. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2019/205 [CORPUS].

[623] Wei Zhang, Andrea Valencia, and Ni-Bin Chang. 2023. Fingerprint Networked Reinforcement Learning via Multiagent Modeling for Improving Decision Making in an Urban Food–Energy–Water Nexus. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2023). doi:10.1109/TSMC.2023.3250620 [CORPUS].

[624] Wenbin Zhang and Jeremy C. Weiss. 2021. Fair Decision-making Under Uncertainty. In *Proceedings of the IEEE International Conference on Data Mining*. doi:10.1109/ICDM51629.2021.00100 [CORPUS].

[625] Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, Longtao Zheng, Xinrun Wang, and Bo An. 2024. A Multimodal Foundation Agent for Financial Trading: Tool-Augmented, Diversified, and Generalist. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. doi:10.1145/3637528.3671801 [CORPUS].

[626] Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. doi:10.1145/3351095.3372852 [CORPUS].

[627] Yiming Zhang, Sravani Nanduri, Liwei Jiang, Tongshuang Wu, and Maarten Sap. 2023. BiasX: "Thinking Slow" in Toxic Content Moderation with Explanations of Implied Social Biases. In *Conference on Empirical Methods in Natural Language Processing*. doi:10.18653/v1/2023.emnlp-main.300 [CORPUS].

[628] Yuan Zhang, Xi Yang, Julie Ivy, and Min Chi. 2019. Attain: attention-based time-aware LSTM networks for disease progression modeling. In *International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2019/607 [CORPUS].

[629] Zaixuan Zhang, Zhansheng Chen, and Liying Xu. 2022. Artificial intelligence and moral dilemmas: perception of ethical decision-making in AI. *Journal of Experimental Social Psychology* (2022). doi:10.1016/j.jesp.2022.104327 [CORPUS].

[630] Guoliang Zhao, Daniel Alencar Costa, and Ying Zou. 2019. Improving the pull requests review process using learning-to-rank algorithms. *Empirical Software Engineering* 24, 6 (2019), 3207–3245. doi:10.1007/s10664-019-09696-8 [CORPUS].

[631] Ming Zhao and Xiang Liu. 2018. Development of decision support tool for optimizing urban emergency rescue facility locations to improve humanitarian logistics management. *Safety Science* (2018). doi:10.1016/j.ssci.2017.10.007 [CORPUS].

[632] Michelle Zhao, Reid Simmons, and Henny Admoni. 2022. The role of adaptation in collective human–AI teaming. *Topics in Cognitive Science* (2022). doi:10.1111/tops.12633 [CORPUS].

[633] Minrui Zhao, Gang Wang, Qiang Fu, Wen Quan, Quan Wen, Xiaoqiang Wang, Tengda Li, Yu Chen, Shan Xue, and Jiaozhi Han. 2024. Intelligent decision-making system of air defense resource allocation via hierarchical reinforcement learning. *International Journal of Intelligent Systems* (2024). doi:10.1155/2024/7777050 [CORPUS].

[634] Shengjia Zhao, Michael Kim, Roshni Sahoo, Tengyu Ma, and Stefano Ermon. 2021. Calibrating predictions to decisions: a novel approach to multi-class calibration. In *Advances in Neural Information Processing Systems*. [CORPUS].

[635] Tianyu Zhao, Mojtaba Taherisadr, and Salma Elmalaki. 2024. FAIRO: Fairness-aware Sequential Decision Making for Human-in-the-Loop CPS. In *ACM/IEEE International Conference on Cyber-Physical Systems*. doi:10.1109/ICCPS61052.2024.00015 [CORPUS].

[636] Chengbo Zheng, Yuheng Wu, Chuhan Shi, Shuai Ma, Jiehui Luo, and Xiaojuan Ma. 2023. Competent but Rigid: Identifying the Gap in Empowering AI to Participate Equally in Group Decision-Making. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*. doi:10.1145/3544548.3581131 [CORPUS].

[637] Yu Zheng, Qianyue Hao, Jingwei Wang, Changzheng Gao, Jinwei Chen, Depeng Jin, and Yong Li. 2024. A Survey of Machine Learning for Urban Decision Making: Applications in Planning, Transportation, and Healthcare. *Comput. Surveys* (2024). doi:10.1145/3695986 .

[638] Yongqing Zheng, Han Yu, Yuliang Shi, Kun Zhang, Shuai Zhen, Lizhen Cui, Cyril Leung, and Chunyan Miao. 2020. PIDS: An Intelligent Electric Power Management Platform. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 08 (Apr. 2020), 13220–13227. doi:10.1609/aaai.v34i08.7027 [CORPUS].

[639] Yongqing Zheng, Han Yu, Kun Zhang, Yuliang Shi, Cyril Leung, and Chunyan Miao. 2019. Intelligent decision support for improving power management. In *Proceedings of the International Joint Conference on Artificial Intelligence*. doi:10.24963/ijcai.2019/965 [CORPUS].

[640] Zetao Zheng, Jie Shao, Shilong Deng, Anjie Zhu, Heng Tao Shen, and Xiaofang Zhou. 2024. Cross-Insight Trader: A Trading Approach Integrating Policies with Diverse Investment Horizons for Portfolio Management. In *IEEE International Conference on Data Engineering*. doi:10.1109/ICDE60146.2024.00356 [CORPUS].

[641] Hao Zhong, Zixuan Yuan, Denghui Zhang, Yi Jiang, Shengming Zhang, and Hui Xiong. 2024. Alphavc: A Reinforcement Learning-based Venture Capital Investment Strategy. In *Hawaii International Conference on System Sciences*. [CORPUS].

[642] Joyce Zhou and Thorsten Joachims. 2023. How to Explain and Justify Almost Any Decision: Potential Pitfalls for Accountability in AI Decision-Making. In *ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)*. doi:10.1145/3593013.3593972 [CORPUS].

[643] Xiaolei Zhu, Xindi Tang, Jiaohong Xie, and Yang Liu. 2023. Dynamic Balancing-Charging Management for Shared Autonomous Electric Vehicle Systems: A Two-Stage Learning-Based Approach. In *IEEE International Conference on Intelligent Transportation Systems*. doi:10.1109/ITSC57777.2023.10422187 [CORPUS].

[644] Yang Zou, Ziwei Wang, Jingsi Huang, Jie Song, and Luo Xu. 2024. Multi-Agent Reinforcement Learning for Mobile Energy Resources Scheduling Amidst Typhoons. *IEEE Transactions on Industry Applications* (2024). doi:10.1109/TIA.2024.3463608 [CORPUS].

[645] Caroline E Zsambok and Gary Klein. 2014. *Naturalistic decision making*. Psychology Press.

[646] Cristina Zuheros, Eugenio Martínez-Cámara, Enrique Herrera-Viedma, and Francisco Herrera. 2021. Sentiment analysis based multi-person multi-criteria decision making methodology using natural language processing and deep learning for smarter decision aid. Case study of restaurant choice using tripadvisor reviews. *Information Fusion* (2021). doi:10.1016/j.inffus.2020.10.019 [CORPUS].

[647] Alexandra Zytek, Dongyu Liu, Rhema Vaithianathan, and Kalyan Veeramachaneni. 2022. Sibyl: Understanding and Addressing the Usability Challenges of Machine Learning In High-Stakes Decision Making. *IEEE Transactions on Visualization and Computer Graphics* (2022). doi:10.1109/TVCG.2021.3114864 [CORPUS].

Appendices for "In the Shadow of Judgment"

- **Appendix A: Model and Performance Details**
  Initial exploration with five ML approaches (SVM, Random Forest, Logistic Regression, LSTM, BERT)
  Performance trends during the screening process (retrospective analysis)
  Terminating performance metrics across binary and multi-class classification
  Per-class performance breakdown and confusion matrices
- **Appendix B: Feature Analysis for the Final SVM**
  Top 30 features contributing to relevant vs. irrelevant classifications
  Coefficient visualizations and interpretations
  Per-paper heat-mapped abstracts showing feature influence
  Examples of false positives, false negatives, and boundary cases
- **Appendix C: Robustness Analysis**
  Breakdown of abstract and title screening by database
  Screening-rescreening reliability assessment
  Estimation of missing papers through stratified sampling
  Bootstrapping analysis to estimate effects of potential miss-outs on code distributions
- **Appendix D: Details of LLM-Assisted Coding Process**
  Comparison of independent coding vs. LLM-assisted approaches
  Inter-rater reliability scores across seven coding dimensions
  Performance metrics (Gwet's AC2, Krippendorff's α, percentage agreement)
  LLM hallucination patterns and mapping strategies
- **Appendix E: Abstract Examples**
  Examples demonstrating AI and human influence coding decisions
  Borderline cases with inclusion/exclusion rationale
  Documented exclusion reasons with examples
- **Appendix F: LLM Prompts**
  Venue rank classification prompt
  Abstract-based coding prompt (decision domains, contribution types, decision factors, AI roles, human roles)
  Full-text coding prompt (decision types, AI influence, human influence)
- **Appendix G: Venue Inclusion Explanation**
  Specific venues included from AAAI
  Venues for ACL Anthology, PMLR, and MLR
  NeurIPS track specifications for 2023–2024
- **Appendix H: List of Supplementary Materials**

A Model Details

A.1 Initial Exploration (original results)

In Sec. 2.4, we ran the first batch of models when we manually screened 17,690 records (44.58% of our corpus). We evaluated five machine learning approaches common for classifying short text. We used TF-IDF vectorization (models 1–3) or neural embeddings (models 4–5) to represent textual content. The corresponding script is included in our supplementary materials (`Analysis/3-label-prediction (Colab).pdf`).

(1) **Support Vector Machine (SVM)**: A linear classifier that finds the optimal boundary separating relevant from irrelevant papers by maximizing the margin between categories in high-dimensional text feature space. We use `kernel='linear', probability=True`, and other hyperparameters are the default of `sklearn.svm.SVC`.
(2) **Random Forest**: An ensemble method combining multiple decision trees, each trained on different subsets of the data, with predictions determined through majority voting across trees. We use `n_estimators=100, random_state=42`, and other hyperparameters are set to the default values of `sklearn.ensemble.RandomForestClassifier`.
(3) **Logistic Regression**: A probabilistic classifier based on the relationship between text features and classes. We use `max_iter=1000, random_state=42`, and other hyper-parameters are the default of `sklearn.linear_model.LogisticRegression`.
(4) **Long Short-Term Memory (LSTM)**: A recurrent neural network architecture processing text sequentially, learning contextual representations through memory cells that selectively retain or discard information across word sequences. We used `tensorflow`, and the model architecture is

```
Embedding(max_words, 128, input_length=max_len),
LSTM(64, dropout=0.2, recurrent_dropout=0.2),
Dense(len(y_unique), activation='softmax')
```

(5) **BERT** – A transformer-based language model pre-trained on large text corpora, fine-tuned on all existing labels to leverage deep bidirectional representations of titles and abstracts for classification. We used the `transformers` package and the pretrained `bert-base-uncased`.

The initial results are below in Tab. 1. For the first three models, we report performance based on 5-fold cross-validation, while we report validation accuracy for the LSTM and BERT.

Table 1: Performance of the five models in our initial exploration

Model	4 classes		2 classes	
	Accuracy ± SD	F1 Macro ± SD	Accuracy ± SD	F1 Macro ± SD
SVM	0.7513 ± 0.0058	0.5035 ± 0.0043	0.7637 ± 0.0075	0.7617 ± 0.0080
Random Forest	0.7497 ± 0.0129	0.4988 ± 0.0107	0.7510 ± 0.0147	0.7525 ± 0.0179
Logistic Regression	0.7554 ± 0.0072	0.5055 ± 0.0052	0.7668 ± 0.0082	0.7632 ± 0.0087
LSTM	0.6938	0.4656	—	—
BERT	0.6323	0.3953	—	—

Given the performance gaps and the training difficulty of LSTM and BERT, we decided to **track SVM, Random Forest, and Logistic Regression**. We also decided to **train on four classes**: Highly relevant, Moderately relevant, Slightly relevant, and Irrelevant, as these labels provide more nuanced classification with lower uncertainty. If we convert 4 classes to 2 classes, it yields much **better accuracy scores**.

A.2 Performance Trend (retrospective results)

We show the performance of SVM, Random Forest, and Logistic Regression during the screening process in Fig. 7. We have to clarify that these results are retrospective: we tracked the accuracy and training data but not the detailed results. So we **retrained all models to replicate the process**. There are very small discrepancies, but the retrospective results are close enough to show the trends. For example, on our last training process, the retrospective results are the same as the original results: 80.80% (accuracy), 40.34% (F1), and 83.86% (Soft accuracy). The SVM has **the best performance**, **a soft accuracy of 84.55%**, when sample size is 13,848.

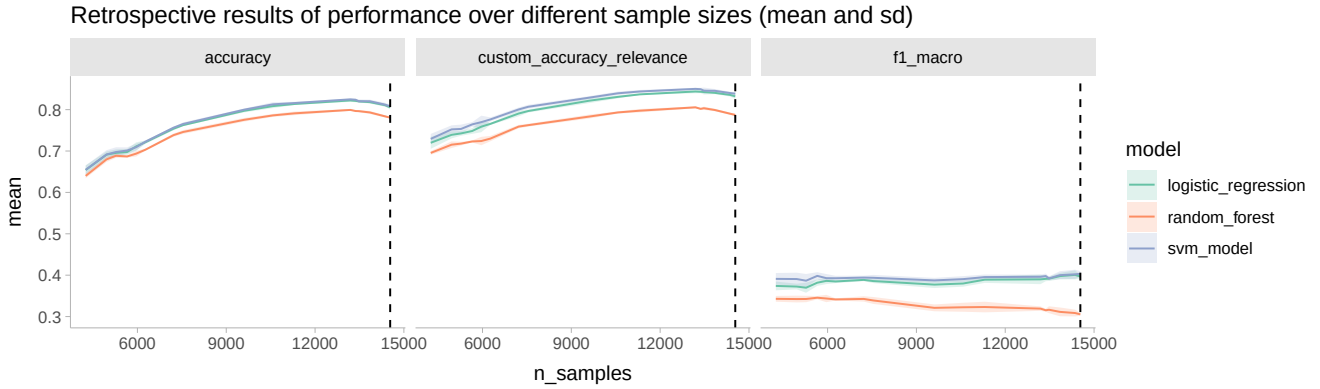


Figure 7: Our retrospective results of performance changes over time and different sample sizes. In the early stage, the performance increases with more screening results; however, accuracies start to drop with F1 being stable. And the uncertainty also reduces. So we think more screening results wouldn't result in better performance.

At this point, the two authors contributed roughly the same number of labels. We decided to terminate the process when we saw **the accuracy scores start to decrease**, and more screening by one author could bias the data towards their results.

A.3 The Terminating Performance (original results)

We show the terminating performance for Relevant (Highly, Moderately, Slightly) vs. Irrelevant, aggregated across the 5 folds in Tab. 2:

Table 2: Performance of the three models at the terminating point

Model	Accuracy ± SD	F1 Macro ± SD	Soft Accuracy ± SD
SVM	**0.8080 ± 0.0045**	**0.4034 ± 0.0064**	**0.8386 ± 0.0028**
Random Forest	0.7812 ± 0.0018	0.3058 ± 0.0025	0.7870 ± 0.0015
Logistic Regression	0.8055 ± 0.0042	0.3966 ± 0.0042	0.8322 ± 0.0064

And the confusion matrices are in Tab. 3:

Table 3: Confusion matrices of Relevant (Highly, Moderately, Slightly) vs. Irrelevant

Model	True Class	Pred Relevant	Pred Irrelevant
SVM	Relevant	**1,406 (9.67%)**	**2,081 (14.32%)**
	Irrelevant	265 (1.82%)	10,785 (74.19%)
Logistic Regression	Relevant	1,256 (8.64%)	2,231 (15.35%)
	Irrelevant	208 (1.43%)	10,842 (74.58%)
Random Forest	Relevant	495 (3.40%)	3,031 (20.85%)
	Irrelevant	**39 (0.27%)**	**11,011 (75.74%)**

As false positives (i.e., Irrelevant misclassified as Relevant) could be detected in the subsequent screening stage, we want to **minimize false negatives** (i.e., missing relevant papers). Therefore, **SVM is the best-performing model**, and we decide to apply it to the remaining papers, with a 14.32% FN rate in mind.

A.4 The Per-Class Terminating Performance (original results)

We show the terminating performance for the per-class performance below, aggregated across the 5 folds in Tab. 4:

Table 4: Per-class performance of the three models at the terminating point

Model	Class	Precision	Recall	F1-Score
SVM	Highly	0.6108	**0.5642**	**0.5866**
	Moderately	0	0	0
	Slightly	0.3964	0.0740	0.1247
	Irrelevant	**0.8383**	**0.9760**	**0.9019**
Random Forest	Highly	**0.7244**	0.2298	0.3489
	Moderately	0	0	0
	Slightly	0.3077	0.0027	0.0054
	Irrelevant	0.7841	0.9965	0.8777
Logistic Regression	Highly	0.6385	0.5007	0.5612
	Moderately	0	0	0
	Slightly	**0.3978**	**0.0754**	**0.1267**
	Irrelevant	0.8293	0.9812	0.8989

And the confusion matrices are in Tab. 5:

Table 5: Confusion matrices of different models (Columns: True Labels, Rows: Predictions)

Model	True Class	Pred Highly	Pred Irrelevant	Pred Moderately	Pred Slightly
SVM	Highly	852 (5.86%)	587 (4.04%)	0	71 (0.48%)
	Irrelevant	191 (1.31%)	10,785 (74.19%)	0	74 (0.51%)
	Moderately	140 (0.96%)	343 (2.35%)	0	21 (0.14%)
	Slightly	212 (1.46%)	1,151 (7.92%)	1 (0.007%)	109 (0.75%)
Random Forest	Highly	347 (2.39%)	1,158 (7.97%)	0	5 (0.034%)
	Irrelevant	36 (0.25%)	11,011 (75.74%)	1 (0.007%)	2 (0.021%)
	Moderately	45 (0.31%)	457 (3.14%)	0	2 (0.014%)
	Slightly	51 (0.35%)	1,416 (9.74%)	2 (0.014%)	4 (0.028%)
Logistic Regression	Highly	756 (5.20%)	683 (4.70%)	1 (0.007%)	70 (0.48%)
	Irrelevant	133 (0.91%)	10,842 (74.58%)	0	75 (0.51%)
	Moderately	129 (0.89%)	352 (2.42%)	0	23 (0.16%)
	Slightly	166 (1.14%)	1,196 (8.22%)	0	111 (0.76%)

Our **takeaways** are:

- SVM is the overall best-performing model, capturing more relevant papers.
- More than half of the FN errors (8% out of 14%) come from distinguishing 'Slightly' from 'Irrelevant'—this was even difficult for us (coders) to distinguish.
- One third of the FN errors (4% out of 14%) come from distinguishing 'Highly' from 'Irrelevant'.

The FN rates are potentially concerning, but we can understand their impacts by looking at the remaining after the sequential screening stage. Also, please note that **trusting FN rates assumes that authors were randomly screening papers**. This is not the case. Combining these two reasons, we will try a different approach to understand the impact of errors in Appx C.

B Feature Analysis for the Final SVM

We also conduct a feature analysis to understand which textual features most influenced the SVM classification. Please see Figs. 8 and 9. We aggregate each feature's standardized coefficients across cross-validation folds, computing both means and standard deviations for the "relevant" and "irrelevant" classes, as well as their difference.

Our takeaways: These coefficients suggest that **the model behavior is largely aligned with our screening process**: positive coefficients tend to surface domain-relevant cues, while unexpected features (such as "participants" or "assisted") can be traced back to corpus artifacts rather than systematic errors. Overall, the feature importance is coherent with our expectations.

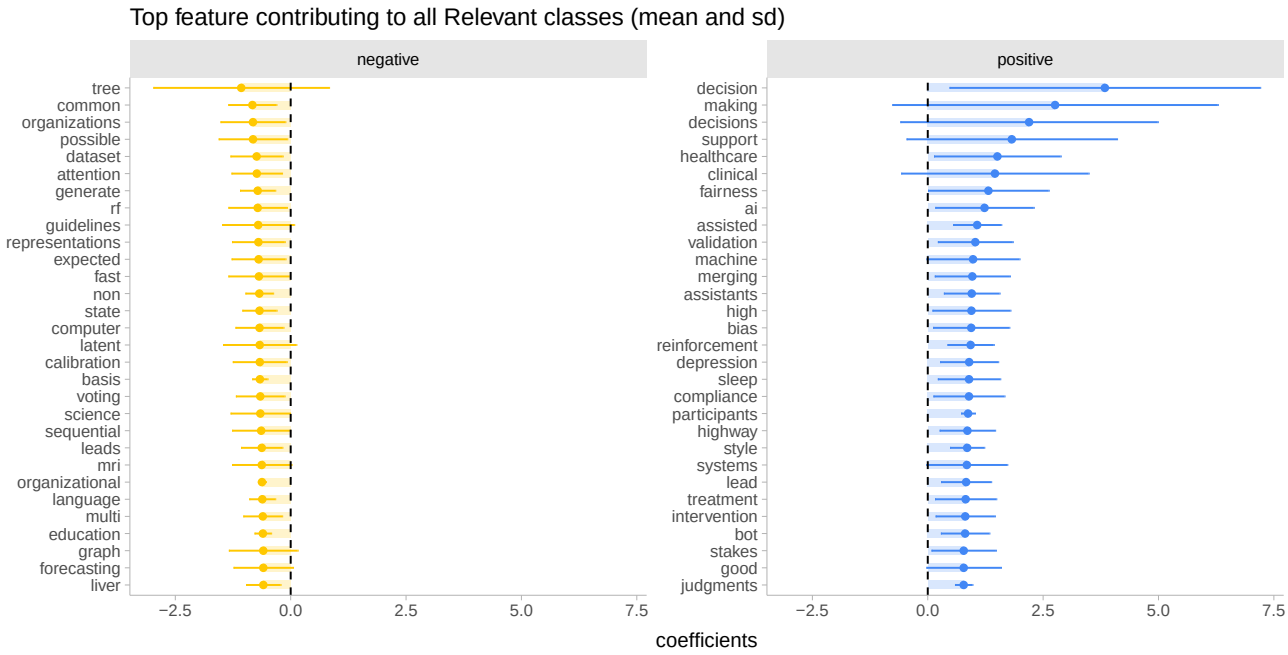


Figure 8: Top 30 features contributing to the all Relevant (Highly, Moderately, Slightly) classes. Words like "decision", "making", "healthcare" and "support" contribute positively, and this matches our intuition. The word "tree" contributes negatively (papers mentioning "decision tree" satisfied searching criteria, but usually will be eliminated if there is no human decision).

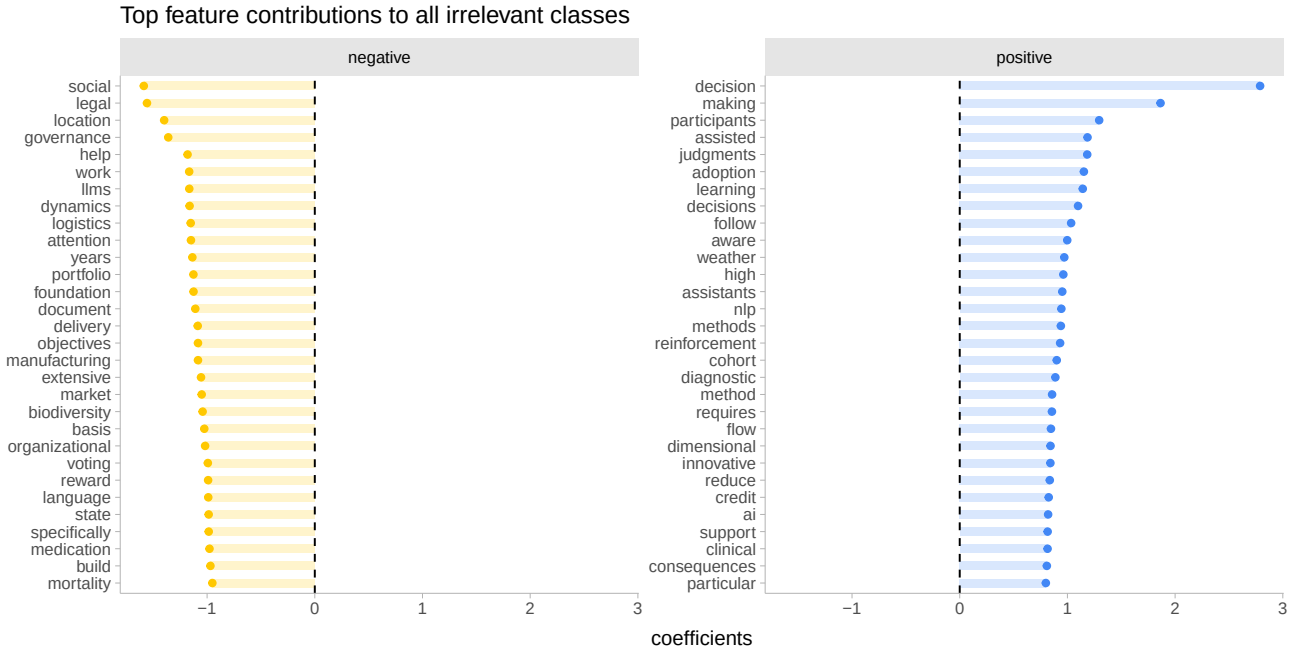


Figure 9: Top 30 features contributing to the Irrelevant class. Words like "decision" and "making" also contribute positively, which also makes sense because many papers mentioning decision are HAID research (e.g., decision tree). The word "participants" is a surprise, but we look into the count of this word, and find many psychology papers mention the word, but don't satisfy other criteria. Similarly, for "assisted", we found many network papers from IEEE contain this word, but they are not HAID.

To complement this aggregate picture, we also generated per-paper visualizations: Figs. 10, 11, 12 and 13. For a small set of manually reviewed examples, we highlighted each token in the abstract with a color corresponding to its **coefficient-difference** (= Relevant − Irrelevant) coefficients. These heat-mapped abstracts illustrate how influential features appear in context. Blue means leaning toward any "Relevant" class, while yellow means leaning toward the "Irrelevant" class. Darker colors mean a bigger absolute value of the coefficient.

Optimization of building demand flexibility using reinforcement learning and rule-based expert systems The increasing use of renewable energy in buildings requires optimization of building demand flexibility to reduce energy costs and carbon emissions. Nevertheless, the optimization process is generally challenging that needs to consider the on-site intermittent energy supply, dynamic building energy demand, and proper utilization of energy storage systems. Leveraging the growing availability of operational data in buildings, data-driven strategies such as reinforcement learning (RL) have emerged as effective approaches to optimizing building demand flexibility. However, training a reliable RL agent is practically data-demanding and time-consuming, limiting its practical applicability. This study proposes a new strategy that integrates a rule-based expert system (RBES) and RL agents into the decision-making process to jointly reduce building energy costs, minimize the peak-to-average ratio (PAR) of grid power, and maximize PV self-consumption. In this strategy, the RBES determines system operation directly in less complex decision-making scenarios, while, in more intricate decision-making environments, it provides a reference decision for RL to explore optimal solutions further. This integration empowers RL agents to avoid unnecessary exploration and significantly enhance learning efficiency. The proposed strategy was tested using PV generation data and energy consumption data of a low energy office building. The results demonstrated an 85.7% improvement in RL learning efficiency and this strategy can successfully avoid sub-optimal convergence during policy learning. Compared to relying solely on the RBES, the proposed strategy led to 5.4% and 19.2% reductions in the electricity costs and daily PAR of grid power at peak hours, respectively. The strategy also achieved a satisfying PV self-consumption ratio of 62.4%, which is merely 0.4% lower than the optimal value determined by the RBES strategy that prioritized maximizing PV self-consumption. Additionally, compared with a model predictive control method developed for cost reduction, the strategy achieved similar cost savings while significantly reducing the decision time.

Figure 10: Coefficient difference of a false negative example: The SVM classifies as "Irrelevant" given many “energy” words, but this is a relevant abstract as it builds an expert system with RL to help make decisions about reducing energy costs. Blue means leaning toward any “Relevant” class, while yellow means leaning toward the “Irrelevant” class.

What type of algorithm is perceived as fairer and more acceptable? A comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs Various types of algorithms are being increasingly used to support public decision-making, yet we do not know how these different algorithm types affect citizens' attitudes and behaviors in specific public affairs. Drawing on public value theory, this study uses a survey experiment to compare the effects of rule-driven versus data-driven algorithmic decision-making (ADM) on citizens' perceived fairness and acceptance. This study also examines the moderating role of familiarity with public affairs and the mediating role of perceived fairness on the relationship. The findings show that rule-driven ADM is generally perceived as fairer and more acceptable than data-driven ADM. Low familiarity with public affairs strengthens citizens' perceived fairness and acceptance of rule-driven ADM more than data-driven ADM, and citizens' perceived fairness plays a significant mediating role in the effect of rule-driven ADM on citizens' acceptance behaviors. These findings further imply that citizens' perceived fairness and acceptance of ADM is strongly shaped by how they perceive familiarity of the decision-making context. In high-familiarity AI application scenarios, the realization of public values may ultimately not be what matters for ADM acceptance among citizens.

Figure 11: Coefficient difference of a Relevant example (correctly classified): this paper is about the perception of algorithmic decision-making. Blue means leaning toward any “Relevant” class, while yellow means leaning toward the “Irrelevant” class.

Emotion Attention-Aware Collaborative Deep Reinforcement Learning for Image Cropping This paper proposes a collaborative deep reinforcement learning model for automatic image cropping (called CDRL-IC). By modeling image cropping as a decision-making process of reinforcement learning, our model could generate optimal cropping result in a few moving and zooming steps. An image with good composition is a comprehensive result by considering the relative importance of objects and also the spatial organization of visual elements. Therefore, emotion attention information which indicates the relationship and importance between objects is applied together with contextual information of color image for image cropping. In order to sufficiently use the emotion attention map and the color image, they are processed by two collaborative agents. The two agents make their primary learning separately and then share information through an information interaction module for making joint action prediction. In order to efficiently evaluate the cropping quality in the reward function, weighted Intersection Over Union (WIoU) is designed by integrating emotion attention map in the traditional IoU. Our CDRL-IC model is tested on a variety of datasets for both image cropping and thumbnail generation. The experiments show that our CDRL-IC model outperforms state-of-the-art methods on these benchmark datasets.

Figure 12: Coefficient difference of an Irrelevant example (correctly classified): This paper is about image processing algorithm and the word "decision-making" is used as a term for an algorithmic process. Blue means leaning toward any "Relevant" class, while yellow means leaning toward the "Irrelevant" class.

Harnessing artificial intelligence for prostate cancer management Summary Prostate cancer (PCa) is a common malignancy in males. The pathology review of PCa is crucial for clinical decision-making, but traditional pathology review is labor intensive and subjective to some extent. Digital pathology and whole-slide imaging enable the application of artificial intelligence (AI) in pathology. This review highlights the success of AI in detecting and grading PCa, predicting patient outcomes, and identifying molecular subtypes. We propose that AI-based methods could collaborate with pathologists to reduce workload and assist clinicians in formulating treatment recommendations. We also introduce the general process and challenges in developing AI pathology models for PCa. Importantly, we summarize publicly available datasets and open-source codes to facilitate the utilization of existing data and the comparison of the performance of different models to improve future studies.

Figure 13: Coefficient difference for a false positive case. There seems to be an HAID process but they also use it merely as motivation for their work. This is classified as "Relevant" but we think this doesn't provide knowledge about HAID. Blue means leaning toward any "Relevant" class, while yellow means leaning toward the "Irrelevant" class.

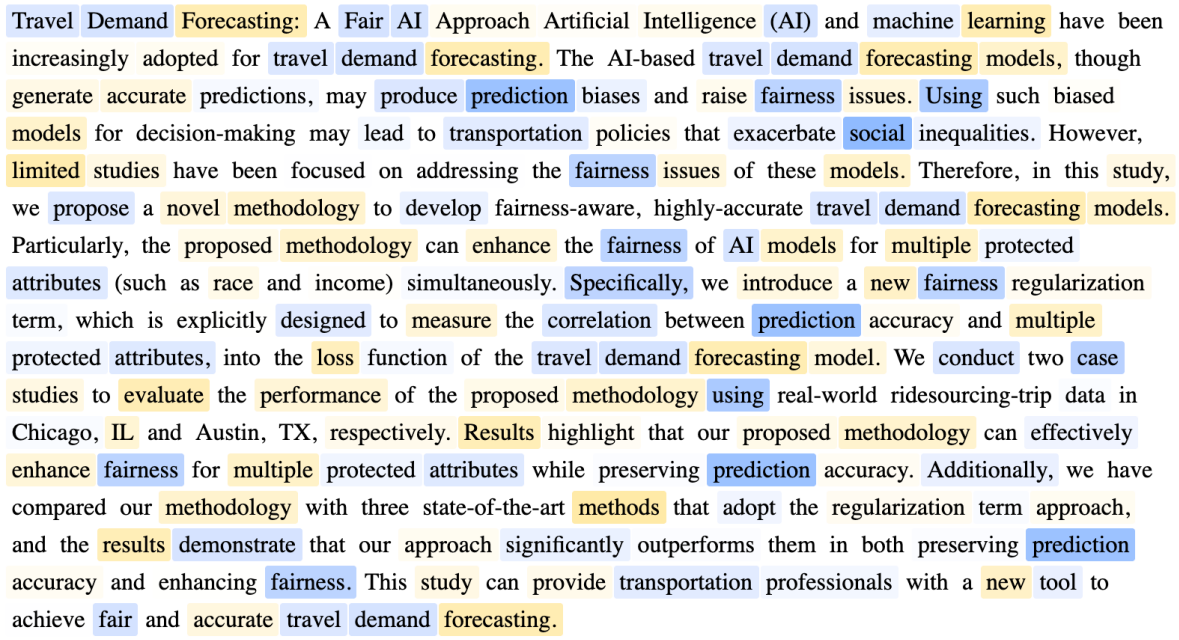


Figure 14: Coefficient difference for a borderline/boundary case. We screened twice, but got different results. They contribute an algorithm for a travel plan, which can be thought of as AI 'executing' a decision, but the connection is weak. The SVM classified this as "Relevant." Blue means leaning toward any "Relevant" class, while yellow means leaning toward the "Irrelevant" class.

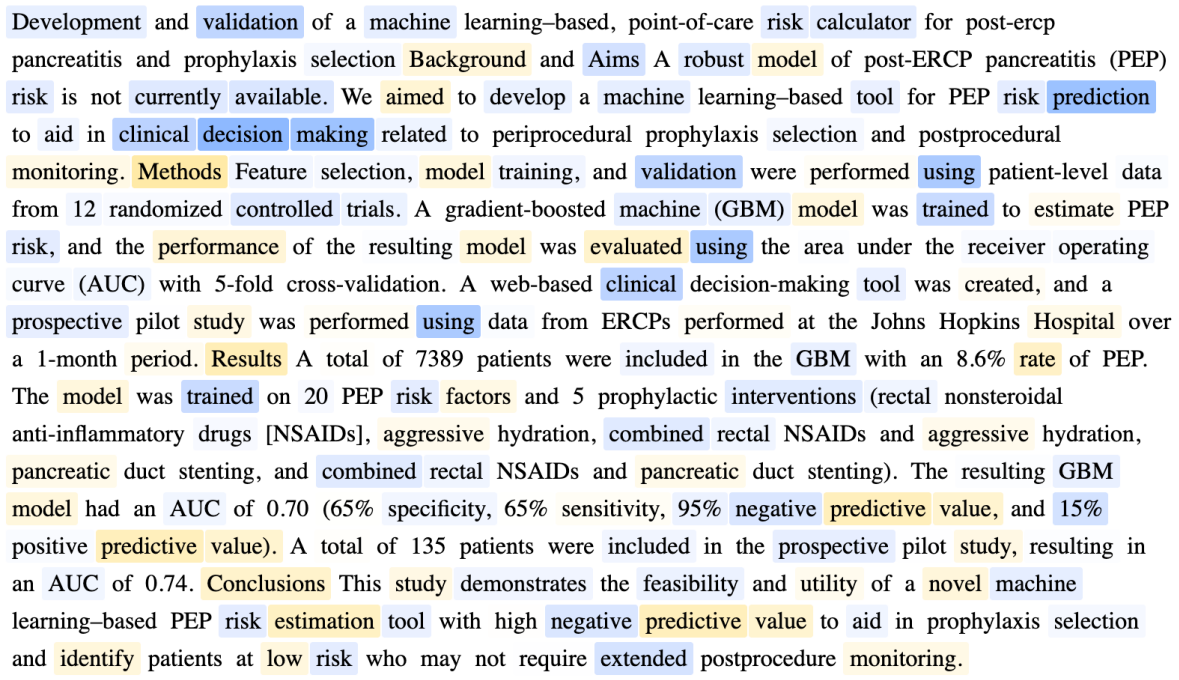


Figure 15: Coefficient difference for a borderline/boundary case. We screened twice, but got different results. They use patient data to build models, and the connection to decision-making is weak and implicit. The SVM classified this as "Relevant." Blue means leaning toward any "Relevant" class, while yellow means leaning toward the "Irrelevant" class.

C Robustness Analysis

The 14.32% FN rate in Iteration 2 does not indicate that we "lost" 1,388 papers (=14.32% * 9690). These predictions were applied only to specific databases, and every flagged item subsequently went through **an additional round** of manual screening (Iteration 3; see Sec. 2.4). Relying on an automated method for a part was reasonable:

(1) Elsevier, Springer Nature, and IEEE contain vast numbers of marginally relevant records (e.g., medical forecasting, autonomous driving), and screening them exhaustively produced fewer new insights.
(2) After multiple rounds of discussion, several borderline cases from other databases/publishers remained genuinely hard to classify. Rather than repeatedly revisiting the same items, we used the model to surface patterns and reduce noise, allowing us to use our manual screening where human judgment was most needed.
(3) We also want to acknowledge a practical but important consideration: manual screening at this scale carries a non-trivial cognitive load. After 10 weeks of intensive review, fatigue affects the consistency and reliability. Using an automated pass allowed us to maintain coder well-being and ensure the quality of the final screening results.

We show a breakdown in Tab. 6. In practice, we prioritize other databases, especially ACM and all AI venues, but also screen at least 15% of each database to ensure representativeness.

Table 6: Breakdowns of the abstract and title screening process

Database	No AI Keywords *(author screened)*	Train Data *(author screened)*	SVM Predicted *(model output)*	No abstract *(authors checked full texts)*
Elsevier	1407	977	5753	63
IEEE	27	2386	2126	0
Springer Nature	9598	703	1682	43
ACM	7	1594	62	0
AAAI	0	1262	32	1
ACL	0	794	27	0
IJCAI	0	723	4	0
APA	1056	29	3	0
AMA	519	157	1	5
ACS	529	238	0	4
AISeL	3	355	0	17
ICLR	0	467	0	0
JSTOR	1884	83	0	58
MLR	0	287	0	0
NeurIPS	3	1512	0	0
PMLR	0	1437	0	0
Sage	16	215	0	3
Science	9	26	0	4
Taylor & Francis	22	320	0	2
Wiley	187	972	0	3

C.1 Estimate of Missing Papers

To estimate how many papers we might have unintentionally excluded, we took a stratified sample of 1,000 papers from the full set of 9,690 (stratified by database). Among these 1000 papers, 863 were classified as 'Irrelevant' and discarded, and 137 papers had been subsequently screened by authors in the final screening stage. We then manually applied our final screening criteria to each sampled paper.

These 863 papers could provide an estimation of the impact of false negatives. Therefore, two primary coders (the same coders as before) judged whether each paper qualifies as HAID research—that is, whether it examines AI used in a decision-making task originally carried out by humans (final screening criteria). Each coder first worked on a subset, then checked the others' decisions, and finally resolved any remaining uncertain cases via offline discussion. The results are shown in Tab. 7 below.

Table 7: Comparison between the original corpus and the manually-screened subset for those 863 papers removed by the SVM

Included in the original corpus	Included in manual screening	N	%	Interpretation
No	No	832	96.41%	no impact
No	Yes	31	3.59%	missing from the corpus
for analyzing human & AI influence (full-text)				
No	No	858	99.42%	no impact
No	Yes	5	0.58%	missing from the corpus

Using the outcomes of this manual screening, we **estimated the number of papers likely missing from the final corpus**. Concretely, for all papers that were marked as 'Irrelevant' by the model (among 9690) , we can use the percentage numbers from our manual screening to estimate how many papers we might have missed, shown in Tab. 8 and Tab. 9 below.

These are just estimates. So we might have missed ~304 papers (~20%) from the final corpus, and ~50 papers (~10%) when analyzing AI and human influences. Typically, qualitative analysis requires only 30% data to generate all codes, so this may influence our code distribution but **shouldn't affect their coverage**. We also **coded the 31 miss-outs** from the sample, and **they are captured by the**

Table 8: Manual screening and miss-out estimates by database

	From manually screening the 863 papers				Estimation of missed papers			
Database	**# in the sample**	**Not include**	**Include**	**Rate**	**SVM removed**	**Est. final total**	**Curr. total**	**Diff**
Elsevier	504	492	12	2.38%	4978	342	223	**-119**
IEEE	202	190	12	5.94%	2004	295	176	**-119**
ACM	5	5	0	0	58	295	295	0
Springer Nature	147	140	7	4.8%	1396	220	154	**-66**
AISeL	0	0	0	0	0	122	122	0
Taylor & Francis	0	0	0	0	0	72	72	0
AAAI	4	4	0	0	31	63	63	0
Wiley	0	0	0	0	0	60	60	0
IJCAI	0	0	0	0	4	54	54	0
NeurIPS	0	0	0	0	0	50	50	0
PMLR	0	0	0	0	0	50	50	0
ACL	1	1	0	0	23	49	49	0
Sage	0	0	0	0	0	30	30	0
ICLR	0	0	0	0	0	19	19	0
AMA	0	0	0	0	1	13	13	0
MLR	0	0	0	0	0	12	12	0
JSTOR	0	0	0	0	0	10	10	0
APA	0	0	0	0	3	7	7	0
ACS	0	0	0	0	0	5	5	0
Science	0	0	0	0	0	4	4	0

Table 9: Manual screening and miss-out estimates for full-text proportions

	From manually screening the 836 papers				Estimation of missed papers in full-texts			
Database	**# in the sample**	**Not include**	**Include**	**%**	**SVM removed**	**Est. final total**	**Curr. total**	**Diff**
ACM	5	5	0	0	58	184	184	0
Elsevier	504	501	3	0.60%	4978	107	77	**-30**
AISeL	0	0	0	0	0	52	52	0
Springer Nature	147	147	0	0	1396	37	37	0
Taylor & Francis	0	0	0	0	0	29	29	0
IEEE	202	200	2	0.99%	2004	31	11	**-20**
ACL	1	1	0	0	23	5	5	0
Wiley	0	0	0	0	0	18	18	0
AAAI	4	4	0	0	31	15	15	0
IJCAI	0	0	0	0	4	13	13	0
Sage	0	0	0	0	0	9	9	0
NeurIPS	0	0	0	0	0	8	8	0
PMLR	0	0	0	0	0	6	6	0
MLR	0	0	0	0	0	4	4	0
AMA	0	0	0	0	1	3	3	0
APA	0	0	0	0	3	3	3	0
ICLR	0	0	0	0	0	3	3	0
ACS	0	0	0	0	0	0	0	0
Science	0	0	0	0	0	0	0	0
JSTOR	0	0	0	0	0	1	1	0

existing codes. Before we give a detailed analysis, let's address screening reliability first.

C.2 Screening Reliability

From the 1000 sampled papers, we screened 137, which were classified as "Relevant"—they had been screened by us before. These papers are usually very marginal, and our screening and re-screening were months apart, so the risk of memorization is low. The results are in Tab. 10 below.

Table 10: Screen-rescreen comparison between the original corpus and samples from the 1k subset (N=137)

Included in the original corpus	Included in manual screening	N	%	Interpretation
No	No	80	58.39%	agreement
No	Yes	10	7.30%	discrepancy
Yes	No	7	5.11%	discrepancy
Yes	Yes	40	29.20%	agreement
for analyzing human & AI influence (full-text)				
No	No	120	87.59%	agreement
No	Yes	7	5.11%	discrepancy
No	No	1	0.73%	discrepancy
No	Yes	9	6.57%	agreement

So the inter-rater reliability for these papers are Gwet's AC2 = 0.77 and Krippendorff's α = 0.73, representing **good reliability**. Also for full-text, we have Gwet's AC2 = 0.93 and Krippendorff's α = 0.66. This distribution is highly skewed, so this represents **excellent reliability**. We also added these irrelevant-to-relevant 10 papers to the corpus.

C.3 Bootstrapping to Estimate the Effects of Miss-outs

So we might have missed ~119, ~119, and ~66 papers from Elsevier, IEEE, and SpringerNature. We can estimate the final code distribution using bootstrapping: we randomly sample the corresponding numbers from the existing corpus, combine them with the updated current corpus, and estimate the code portion. There are more rigorous Bayesian approaches. Here, we try the simplest approximation to just get a sense of uncertainty. Below is the R code to do so.

```
all_codes %>%
   filter(database == 'ieee') %>%
   sample_n(119, replace = TRUE) %>%
   rbind(all_codes %>%
           filter(database == 'elsevier') %>%
           sample_n(119, replace = TRUE)) %>%
   rbind(all_codes %>%
           filter(database == 'springernature') %>%
           sample_n(66, replace = TRUE)) %>%
   rbind(final_results)
```

Because we also have screening uncertainty when estimating the number of missing papers, we input bounds (±12% from C.2, 105–133, 58–74) to get a sense of the range and repeat this process 500 times for each. We show and discuss their impacts in Figs. 16, 17, 18, & 19.

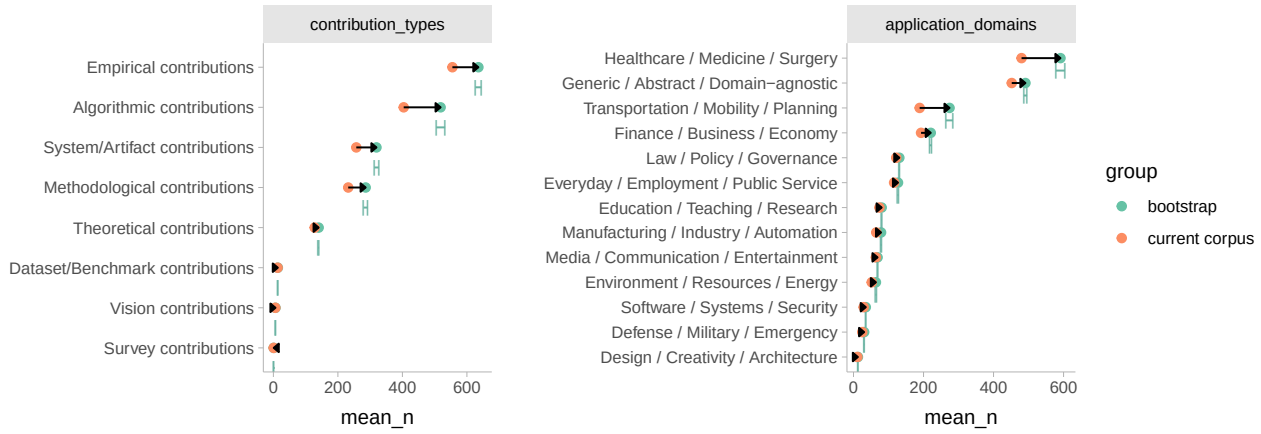


Figure 16: Bootstrap to fill in missing papers vs. the current corpus. So there are potential rank flips, but overall, these ranks are similar. Transportation may exceed Finance. Error bars are results with ± 12% papers. We also draw SDs, but they are too small to see.

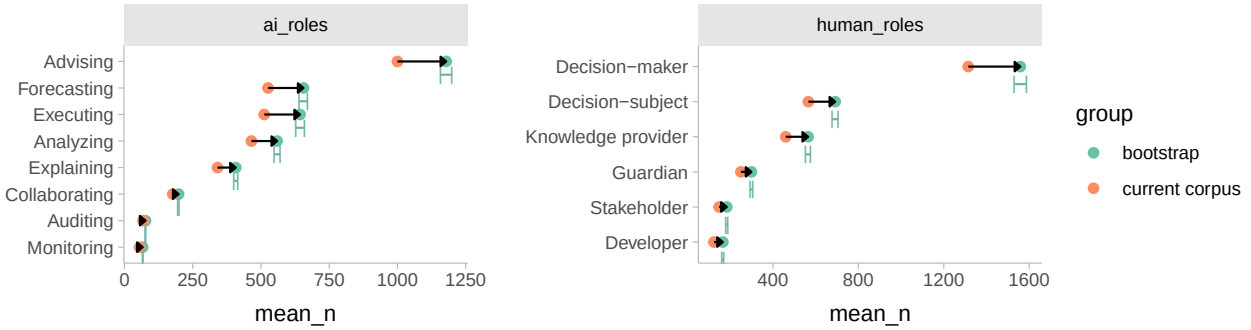


Figure 17: Bootstrap to fill in missing papers vs. the current corpus. Overall, the absolute numbers may change, but the ranks are the same. Error bars are results with ± 12% papers. We also draw SDs, but they are too small to see.

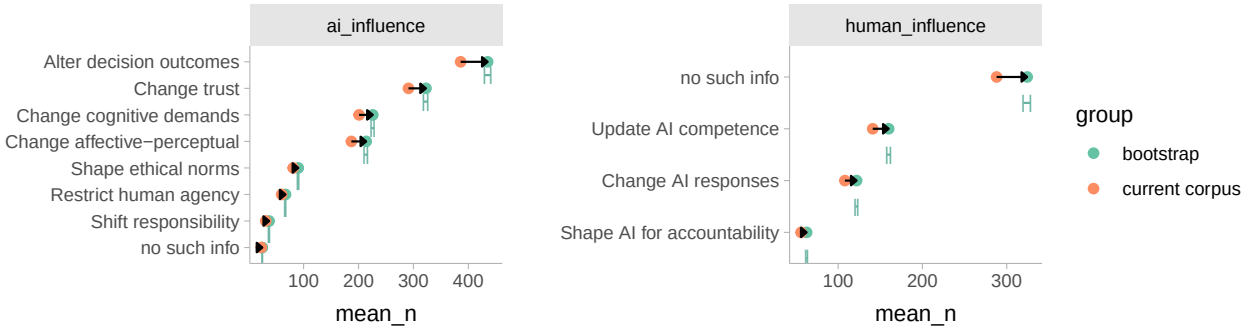


Figure 18: Bootstrap to fill in missing papers vs. the current corpus.. Overall, the absolute numbers may change, but the ranks are the same. Error bars are results with ± 12% papers. We also draw SDs, but they are too small to see.

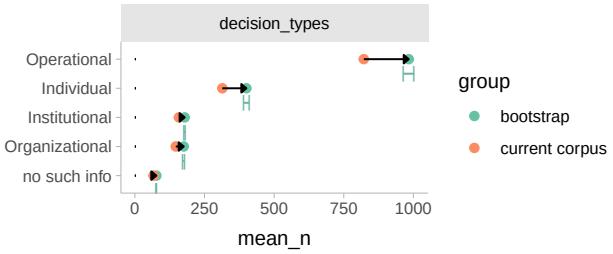


Figure 19: Bootstrap to fill in missing papers vs. the current corpus. Overall, the absolute numbers may change, but the ranks are the same. Error bars are results with ± 12% papers. We also draw SDs but they are too small to see.

D Details of LLM-assisted Coding Process

In our coding process, besides using primary coders, our secondary coders checked all codes with GPT's and Claude's results. To validate these approaches, we sampled 10% (146) papers from the corpus (before we added recovered papers) and applied four approaches:

- a secondary coder acts as an independent coder (`Ind`), and then reconciles disagreement with primary coders (`IndPrim`)
- a secondary coder checking primary coders' initial results by referring to GPT coding results (`wGPT`)
- a secondary coder checking primary coders' initial results by referring to Claude coding results (`wClaude`)
- a secondary coder checking primary coders' initial results by referring to both GPT and Claude coding results (`wGPTClaude`)

We view the human codes after reconciling as "gold standard" (reference) and compare others with this reference using inter-rate reliability (IRR) scores. Because Krippendorff's α and similar metrics are unreliable for skewed distributions like ours [445, 556], we also report Gwet's AC2 [445] and percentage agreement (PA). Also, because multi-label reliability calculation remains an open problem [363], we computed per-label scores and averaged all labels for a dimension.

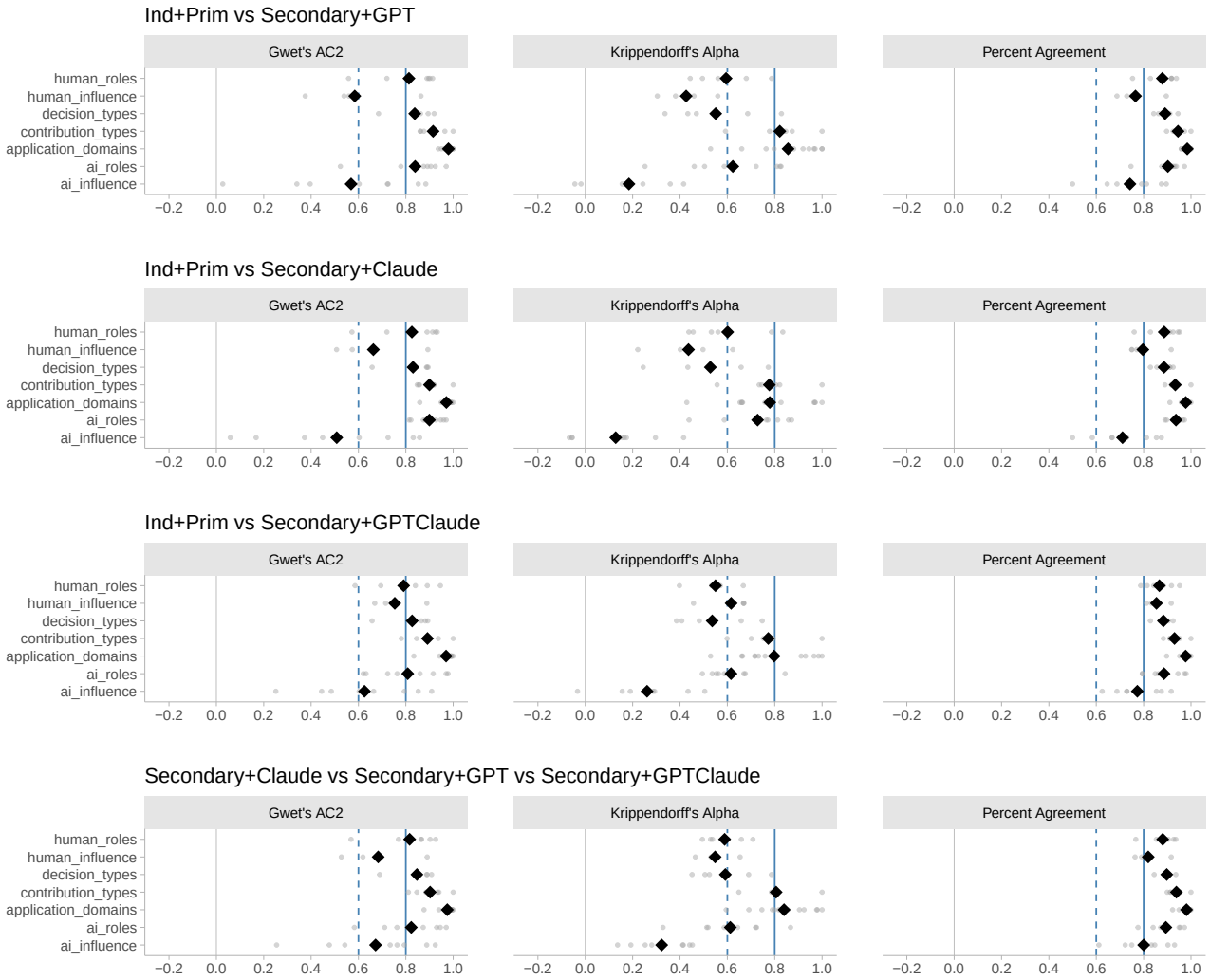


Figure 20: IRR scores for the subset: comparing three LLM-assisted approaches with reference. Overall, the secondary coder working with both LLMs' outputs has better agreement with the reference. However, *AI influence* shows low agreement—we need better instructions. Each small dot (•) represents one label for that dimension, and diamonds (♦) represent mean scores aggregating all labels.

Table 11: IRRs between reference standard and LLM-assisted coders across coding dimensions. These are mean ± SD for Fig. 20. The performance among the three LLM-assisted coders is similar. The two coders with Claude were better than the coder with GPT alone. However, we decided to show both GPT and Claude as the worst dimension (AI influence) is the best under this consideration.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Ind+Prim vs 2nd+GPT	0.57 ± 0.29	0.84 ± 0.14	**0.98 ± 0.02**	0.92 ± 0.06	**0.84 ± 0.09**	0.58 ± 0.20	0.81 ± 0.14
Gwet's AC2	Ind+Prim vs 2nd+Claude	0.51 ± 0.30	**0.90 ± 0.06**	0.97 ± 0.04	0.90 ± 0.06	0.83 ± 0.10	0.66 ± 0.17	**0.83 ± 0.15**
Gwet's AC2	Ind+Prim vs 2nd+Both	**0.63 ± 0.23**	0.81 ± 0.14	0.97 ± 0.04	0.89 ± 0.08	0.83 ± 0.10	**0.75 ± 0.10**	0.79 ± 0.13
Gwet's AC2	2nd+Claude vs 2nd+GPT vs 2nd+Both	0.67 ± 0.23	0.82 ± 0.13	0.98 ± 0.03	0.90 ± 0.07	0.85 ± 0.09	0.68 ± 0.15	0.82 ± 0.13
Krippendorff's α	Ind+Prim vs 2nd+GPT	0.18 ± 0.16	0.62 ± 0.21	0.86 ± 0.14	0.82 ± 0.13	0.55 ± 0.20	0.43 ± 0.11	0.59 ± 0.13
Krippendorff's α	Ind+Prim vs 2nd+Claude	0.13 ± 0.18	0.73 ± 0.15	0.78 ± 0.17	0.78 ± 0.14	0.53 ± 0.20	0.44 ± 0.17	0.60 ± 0.17
Krippendorff's α	Ind+Prim vs 2nd+GPTClaude	0.26 ± 0.17	0.62 ± 0.11	0.80 ± 0.15	0.77 ± 0.13	0.54 ± 0.16	0.62 ± 0.11	0.55 ± 0.09
Krippendorff's α	2nd+Claude vs 2nd+GPT vs 2nd+Both	0.32 ± 0.12	0.61 ± 0.16	0.84 ± 0.12	0.81 ± 0.11	0.59 ± 0.14	0.55 ± 0.08	0.59 ± 0.08
% Agreement	Ind+Prim vs 2nd+GPT	0.74 ± 0.13	0.90 ± 0.07	0.98 ± 0.01	0.95 ± 0.04	0.89 ± 0.04	0.77 ± 0.09	0.88 ± 0.07
% Agreement	Ind+Prim vs 2nd+Claude	0.71 ± 0.13	0.94 ± 0.03	0.98 ± 0.02	0.93 ± 0.04	0.89 ± 0.04	0.80 ± 0.08	0.89 ± 0.08
% Agreement	Ind+Prim vs 2nd+Both	0.77 ± 0.10	0.89 ± 0.07	0.98 ± 0.03	0.93 ± 0.04	0.88 ± 0.04	0.85 ± 0.05	0.87 ± 0.06
% Agreement	2nd+Claude vs 2nd+GPT vs 2nd+Both	0.80 ± 0.10	0.89 ± 0.07	0.98 ± 0.02	0.94 ± 0.04	0.90 ± 0.03	0.82 ± 0.07	0.88 ± 0.06

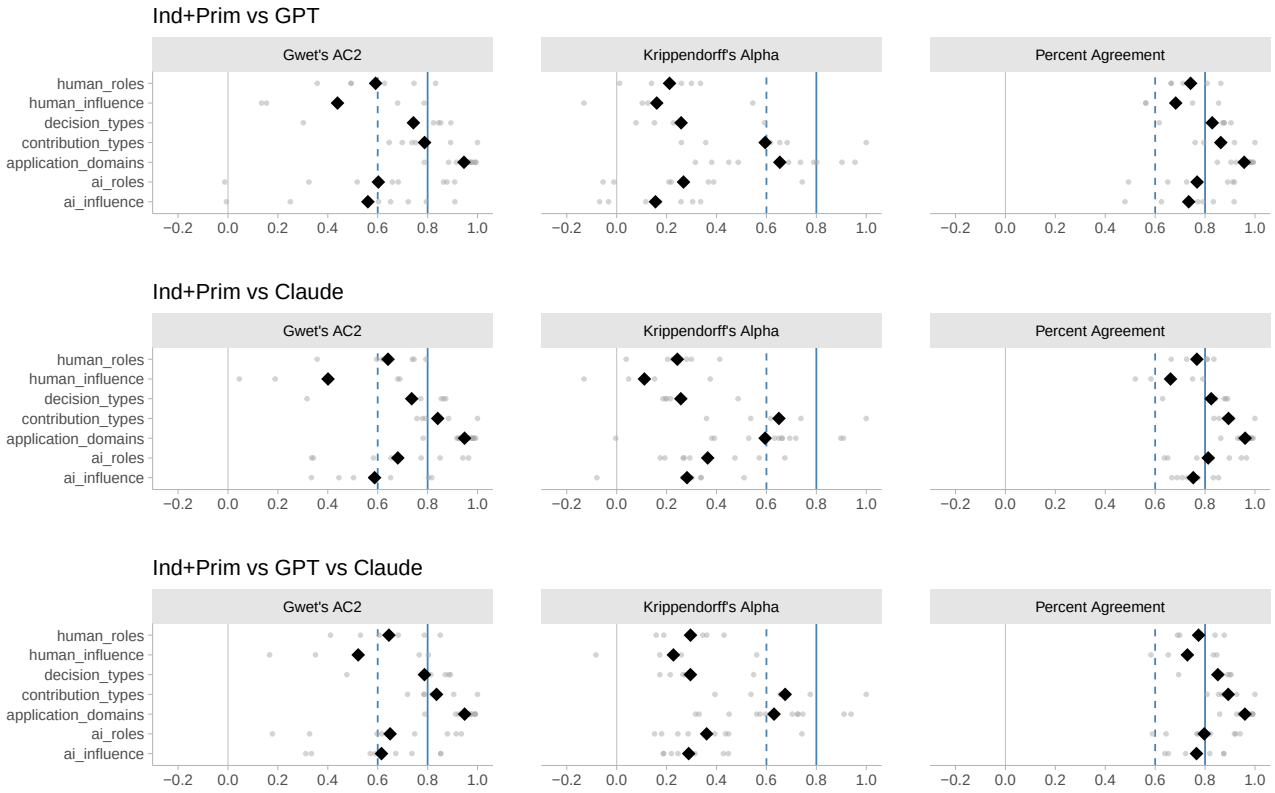


Figure 21: IRR scores for the subset: comparing LLM outputs with"gold standard" (reference). The trends are similar to LLM-assisted approaches. LLMs have very good performance on *application domains* and *contribution types*, and good performance on *human roles*, *AI roles*, and *decision types*. However, LLMs perform very badly on *human influence* and *AI influence*. Also, LLMs introduce higher uncertainty between different labels compared to those with human verification. Each small dot (•) represents one label for that dimension, and diamonds (⧫) represent mean scores aggregating all labels.

Table 12: IRRs between "gold standard" and LLMs. These are mean ± SD for Fig. 21. Claude generally performs better than GPT. This difference between table and Tab. 11 shows the value of human verification.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Ind+Prim vs GPT	0.56 ± 0.30	0.60 ± 0.32	**0.95 ± 0.06**	0.79 ± 0.13	0.74 ± 0.25	**0.44 ± 0.34**	0.59 ± 0.18
Gwet's AC2	Ind+Prim vs Claude	**0.59 ± 0.17**	**0.68 ± 0.25**	**0.95 ± 0.06**	**0.84 ± 0.09**	**0.74 ± 0.24**	0.40 ± 0.33	**0.64 ± 0.16**
Gwet's AC2	Ind+Prim vs GPT vs Claude	0.62 ± 0.21	0.65 ± 0.28	0.95 ± 0.05	0.84 ± 0.10	0.79 ± 0.18	0.52 ± 0.31	0.64 ± 0.16
Krippendorff's α	Ind+Prim vs GPT	0.16 ± 0.15	0.27 ± 0.25	0.65 ± 0.20	0.59 ± 0.26	0.26 ± 0.20	0.16 ± 0.28	0.21 ± 0.12
Krippendorff's α	Ind+Prim vs Claude	0.28 ± 0.16	0.36 ± 0.19	0.59 ± 0.24	0.65 ± 0.21	0.26 ± 0.13	0.11 ± 0.21	0.24 ± 0.12
Krippendorff's α	Ind+Prim vs GPT vs Claude	0.29 ± 0.10	0.36 ± 0.19	0.63 ± 0.19	0.67 ± 0.21	0.30 ± 0.15	0.23 ± 0.27	0.30 ± 0.10
% Agreement	Ind+Prim vs GPT	0.73 ± 0.13	0.77 ± 0.15	0.96 ± 0.04	0.86 ± 0.09	0.83 ± 0.12	0.68 ± 0.14	0.74 ± 0.08
% Agreement	Ind+Prim vs Claude	0.75 ± 0.07	0.81 ± 0.12	0.96 ± 0.03	0.89 ± 0.06	0.82 ± 0.11	0.66 ± 0.13	0.77 ± 0.06
% Agreement	Ind+Prim vs GPT vs Claude	0.77 ± 0.09	0.80 ± 0.13	0.96 ± 0.04	0.89 ± 0.07	0.85 ± 0.09	0.73 ± 0.13	0.77 ± 0.07

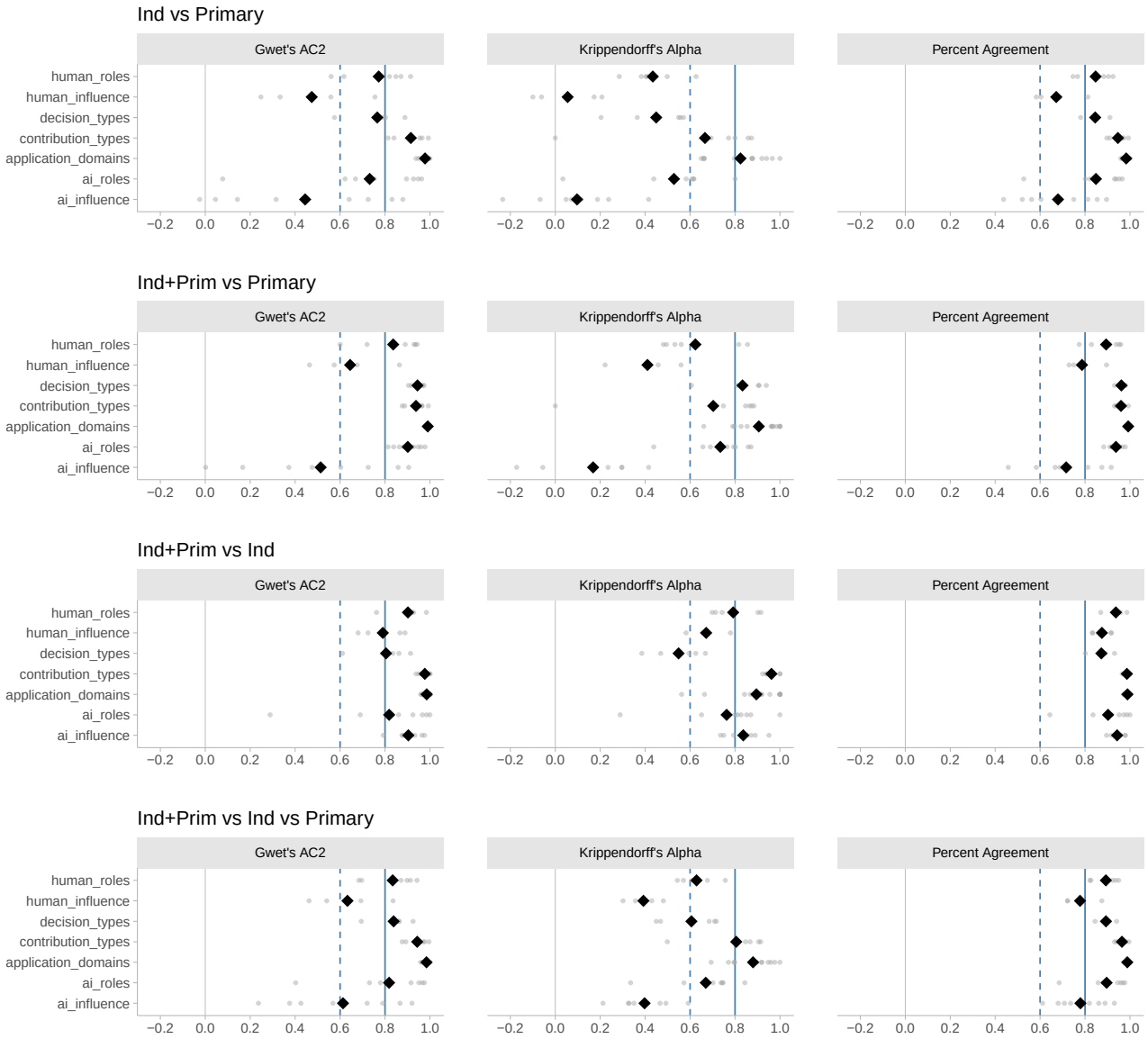


Figure 22: IRR scores for the paper subset: comparing two human coders. The two coders disagree on *AI influence* and *Human influence*. We then prepared more instructions for these two dimensions. Overall, they still have good agreement. Each small dot (•) represents one label for that dimension, and diamonds (♦) represent mean scores aggregating all labels.

Table 13: IRR scores when comparing two human coders. These are mean ± SD for Fig. 22.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Ind vs Prim	0.44 ± 0.37	0.73 ± 0.30	0.98 ± 0.02	0.91 ± 0.07	0.77 ± 0.12	0.47 ± 0.23	0.77 ± 0.15
Gwet's AC2	Ind+Prim vs Prim	0.51 ± 0.32	0.90 ± 0.06	0.99 ± 0.01	0.94 ± 0.05	0.94 ± 0.03	0.64 ± 0.17	0.84 ± 0.14
Gwet's AC2	Ind+Prim vs Ind	0.90 ± 0.06	0.82 ± 0.24	0.99 ± 0.02	0.98 ± 0.03	0.80 ± 0.12	0.79 ± 0.10	0.90 ± 0.07
Gwet's AC2	Ind+Prim vs Ind vs Prim	0.61 ± 0.25	0.82 ± 0.19	0.98 ± 0.01	0.94 ± 0.05	0.84 ± 0.09	0.63 ± 0.17	0.83 ± 0.11
Krippendorff's α	Ind vs Prim	0.10 ± 0.20	0.53 ± 0.22	0.82 ± 0.13	0.67 ± 0.33	0.45 ± 0.16	0.06 ± 0.16	0.43 ± 0.12
Krippendorff's α	Ind+Prim vs Prim	0.17 ± 0.19	0.73 ± 0.14	0.91 ± 0.11	0.70 ± 0.35	0.83 ± 0.14	0.41 ± 0.14	0.62 ± 0.17
Krippendorff's α	Ind+Prim vs Ind	0.84 ± 0.07	0.76 ± 0.21	0.90 ± 0.14	0.96 ± 0.03	0.55 ± 0.12	0.67 ± 0.08	0.79 ± 0.09
Krippendorff's α	Ind+Prim vs Ind vs Prim	0.40 ± 0.12	0.67 ± 0.16	0.88 ± 0.09	0.81 ± 0.16	0.61 ± 0.13	0.39 ± 0.08	0.63 ± 0.08
% Agreement	Ind vs Prim	0.68 ± 0.17	0.85 ± 0.14	0.98 ± 0.01	0.95 ± 0.04	0.85 ± 0.05	0.67 ± 0.10	0.85 ± 0.07
% Agreement	Ind+Prim vs Prim	0.72 ± 0.15	0.94 ± 0.03	0.99 ± 0.01	0.96 ± 0.02	0.96 ± 0.02	0.79 ± 0.07	0.89 ± 0.08
% Agreement	Ind+Prim vs Ind	0.94 ± 0.03	0.90 ± 0.12	0.99 ± 0.01	0.99 ± 0.01	0.87 ± 0.05	0.88 ± 0.05	0.94 ± 0.04
% Agreement	Ind+Prim vs Ind vs Prim	0.78 ± 0.11	0.90 ± 0.10	0.99 ± 0.01	0.96 ± 0.02	0.89 ± 0.03	0.78 ± 0.07	0.89 ± 0.06

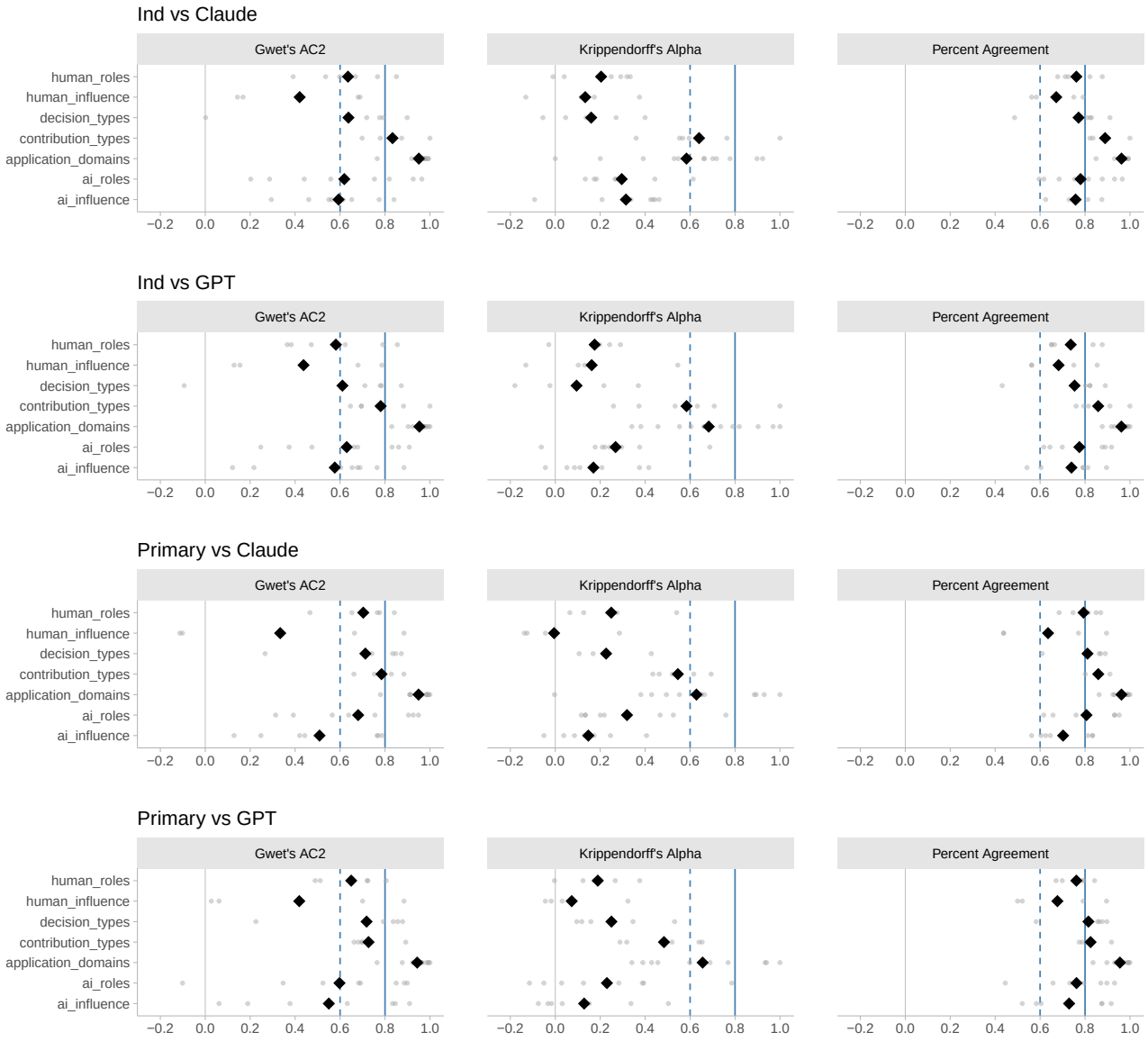


Figure 23: IRR scores for the subset: comparing human coders with LLMs. These show disagreement and variance, which we use to prepare for more instructions. Each small dot (•) represents one label for that dimension, and diamonds (♦) represent mean scores aggregating all labels.

Table 14: IRRs between human coders and LLMs across coding dimensions. These are mean ± SD for Fig. 23.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Ind vs Claude	0.59 ± 0.17	0.62 ± 0.29	0.95 ± 0.06	0.83 ± 0.10	0.64 ± 0.36	0.42 ± 0.31	0.64 ± 0.16
Gwet's AC2	Ind vs GPT	0.58 ± 0.27	0.63 ± 0.24	0.95 ± 0.05	0.78 ± 0.14	0.61 ± 0.40	0.44 ± 0.34	0.58 ± 0.21
Gwet's AC2	Prim vs Claude	0.51 ± 0.27	0.68 ± 0.25	0.95 ± 0.06	0.78 ± 0.08	0.71 ± 0.25	0.33 ± 0.52	0.70 ± 0.13
Gwet's AC2	Prim vs GPT	0.55 ± 0.34	0.60 ± 0.34	0.94 ± 0.06	0.73 ± 0.09	0.72 ± 0.28	0.42 ± 0.44	0.65 ± 0.13
Krippendorff's α	Ind vs Claude	0.31 ± 0.19	0.30 ± 0.16	0.58 ± 0.26	0.64 ± 0.22	0.16 ± 0.18	0.13 ± 0.21	0.20 ± 0.15
Krippendorff's α	Ind vs GPT	0.17 ± 0.16	0.27 ± 0.21	0.68 ± 0.21	0.58 ± 0.26	0.09 ± 0.21	0.16 ± 0.28	0.18 ± 0.11
Krippendorff's α	Prim vs Claude	0.15 ± 0.15	0.32 ± 0.24	0.63 ± 0.27	0.55 ± 0.11	0.23 ± 0.12	-0.01 ± 0.20	0.25 ± 0.16
Krippendorff's α	Prim vs GPT	0.13 ± 0.22	0.23 ± 0.30	0.66 ± 0.22	0.48 ± 0.17	0.25 ± 0.19	0.07 ± 0.17	0.19 ± 0.13
% Agreement	Ind vs Claude	0.76 ± 0.07	0.78 ± 0.14	0.96 ± 0.04	0.89 ± 0.06	0.77 ± 0.16	0.67 ± 0.12	0.76 ± 0.07
% Agreement	Ind vs GPT	0.74 ± 0.12	0.77 ± 0.11	0.96 ± 0.03	0.86 ± 0.09	0.75 ± 0.18	0.68 ± 0.14	0.74 ± 0.10
% Agreement	Prim vs Claude	0.70 ± 0.12	0.81 ± 0.13	0.96 ± 0.04	0.86 ± 0.04	0.81 ± 0.12	0.64 ± 0.23	0.79 ± 0.07
% Agreement	Prim vs GPT	0.73 ± 0.16	0.76 ± 0.16	0.96 ± 0.05	0.82 ± 0.06	0.82 ± 0.13	0.68 ± 0.20	0.76 ± 0.06

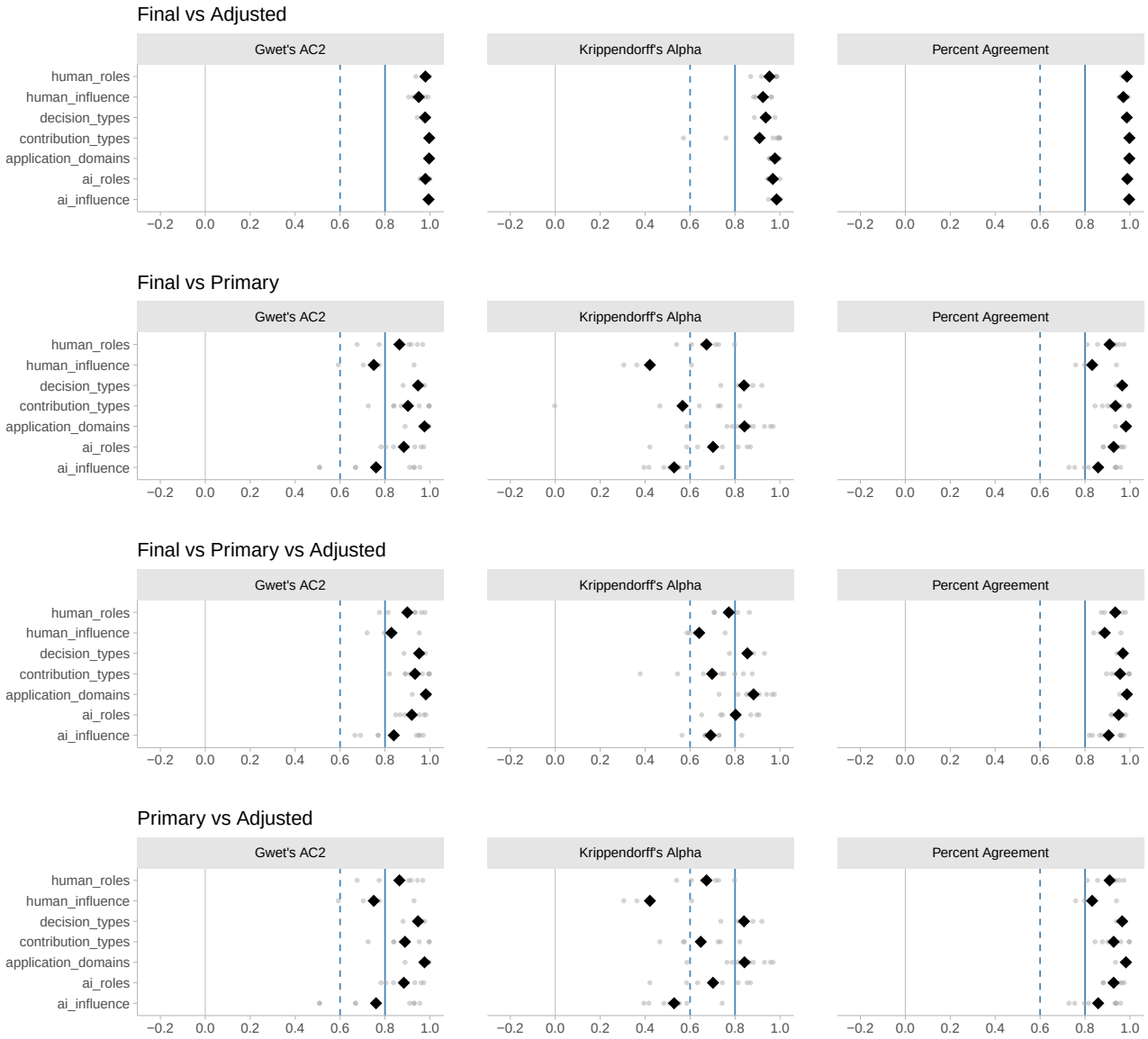


Figure 24: IRR scores for the remaining records (N=1351): comparing final coding results (refined by primary coders) with primary coders' initial results (first) and secondary coders' results (second). The refinements were small, while the final codes show substantial agreement with the primary coders' initial results. Each small dot (•) represents one label for that dimension, and diamonds (⧫) represent mean scores aggregating all labels.

Table 15: IRRs between final, primary, and pre-adjusted codes across all coding dimensions. Mean ± SD for Fig. 24.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Final vs Adjusted	0.99 ± 0.01	0.98 ± 0.02	1.00 ± 0.00	1.00 ± 0.00	0.98 ± 0.02	0.95 ± 0.04	0.98 ± 0.02
Gwet's AC2	Final vs Primary	0.76 ± 0.19	0.88 ± 0.07	0.98 ± 0.03	0.90 ± 0.10	0.95 ± 0.04	0.75 ± 0.14	0.86 ± 0.11
Gwet's AC2	Final vs Primary vs Adjusted	0.84 ± 0.13	0.92 ± 0.05	0.98 ± 0.02	0.93 ± 0.07	0.95 ± 0.04	0.83 ± 0.10	0.90 ± 0.08
Gwet's AC2	Primary vs Adjusted	0.76 ± 0.19	0.88 ± 0.07	0.98 ± 0.03	0.89 ± 0.10	0.95 ± 0.04	0.75 ± 0.14	0.86 ± 0.11
Krippendorff's α	Final vs Adjusted	0.98 ± 0.02	0.97 ± 0.02	0.98 ± 0.01	0.91 ± 0.16	0.94 ± 0.03	0.92 ± 0.04	0.95 ± 0.05
Krippendorff's α	Final vs Primary	0.53 ± 0.11	0.70 ± 0.15	0.84 ± 0.10	0.57 ± 0.26	0.84 ± 0.07	0.42 ± 0.13	0.67 ± 0.09
Krippendorff's α	Final vs Primary vs Adjusted	0.69 ± 0.08	0.80 ± 0.09	0.88 ± 0.06	0.70 ± 0.17	0.85 ± 0.06	0.64 ± 0.08	0.77 ± 0.06
Krippendorff's α	Primary vs Adjusted	0.53 ± 0.11	0.70 ± 0.15	0.84 ± 0.10	0.65 ± 0.12	0.84 ± 0.07	0.42 ± 0.13	0.67 ± 0.09
% Agreement	Final vs Adjusted	1.00 ± 0.00	0.99 ± 0.01	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.01	0.97 ± 0.02	0.99 ± 0.01
% Agreement	Final vs Primary	0.86 ± 0.09	0.93 ± 0.04	0.98 ± 0.02	0.94 ± 0.06	0.97 ± 0.02	0.83 ± 0.08	0.91 ± 0.06
% Agreement	Final vs Primary vs Adjusted	0.90 ± 0.06	0.95 ± 0.02	0.99 ± 0.01	0.96 ± 0.04	0.97 ± 0.02	0.89 ± 0.05	0.93 ± 0.04
% Agreement	Primary vs Adjusted	0.86 ± 0.09	0.93 ± 0.04	0.98 ± 0.02	0.93 ± 0.06	0.97 ± 0.02	0.83 ± 0.08	0.91 ± 0.06

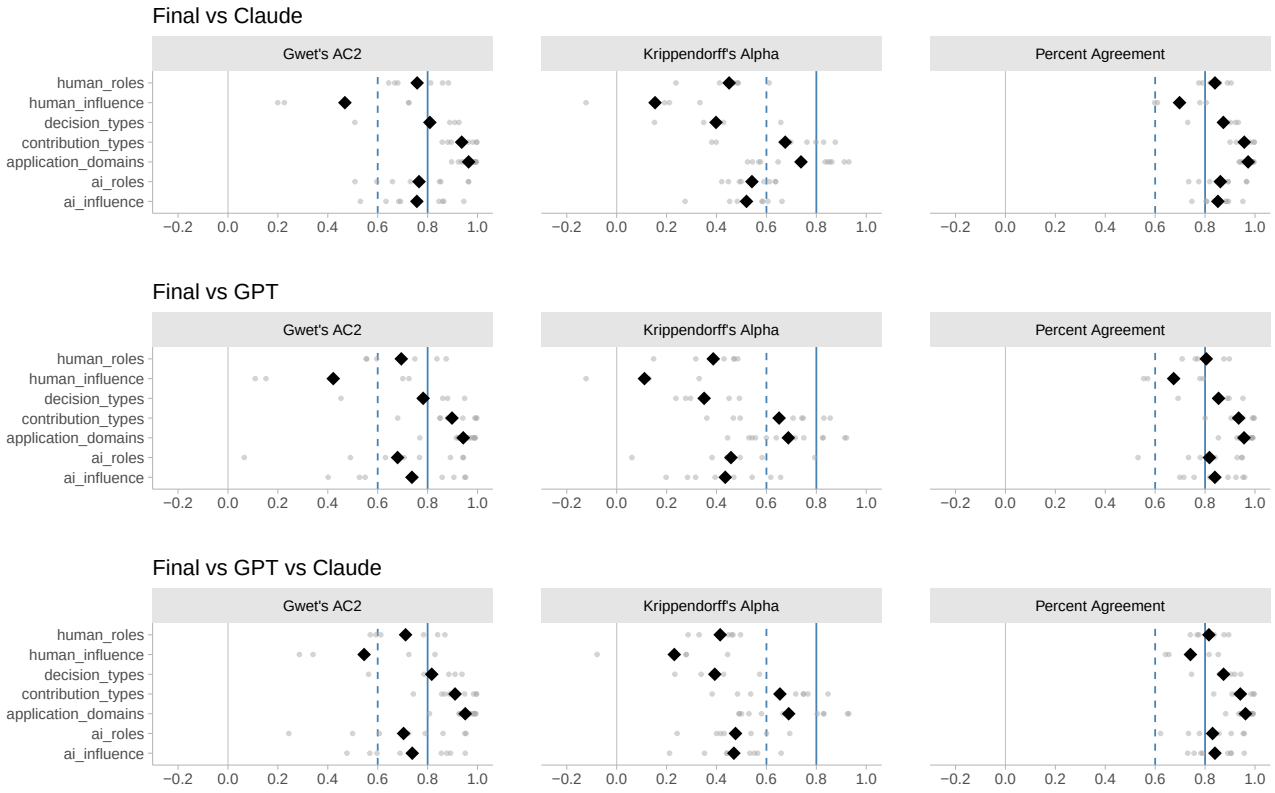


Figure 25: IRR scores for the remaining records (N=1351): comparing final coding results with LLM outputs. LLMs perform well on several dimensions, and Claude performs better than GPT. But both have substantial disagreement with our results on *human influence*. Each small dot (•) represents one label for that dimension, and diamonds (⧫) represent mean scores aggregating all labels.

Table 16: IRRs between final human-coded labels and LLM-assisted labels across all coding dimensions. Mean ± SD for Fig. 25.

Metric	Coders	AI Infl.	AI Roles	Domains	Contrib.	Decision	Human Infl.	Human Roles
Gwet's AC2	Final vs Claude	0.76 ± 0.14	0.77 ± 0.17	0.96 ± 0.03	0.94 ± 0.05	0.81 ± 0.17	0.47 ± 0.30	0.76 ± 0.11
Gwet's AC2	Final vs GPT	0.74 ± 0.22	0.68 ± 0.29	0.94 ± 0.06	0.90 ± 0.11	0.78 ± 0.19	0.42 ± 0.34	0.69 ± 0.14
Gwet's AC2	Final vs GPT vs Claude	0.74 ± 0.18	0.70 ± 0.24	0.95 ± 0.05	0.91 ± 0.09	0.82 ± 0.15	0.55 ± 0.27	0.71 ± 0.13
Krippendorff's α	Final vs Claude	0.52 ± 0.12	0.54 ± 0.09	0.74 ± 0.15	0.68 ± 0.19	0.40 ± 0.18	0.15 ± 0.19	0.45 ± 0.12
Krippendorff's α	Final vs GPT	0.43 ± 0.16	0.46 ± 0.20	0.69 ± 0.15	0.65 ± 0.18	0.35 ± 0.11	0.11 ± 0.19	0.39 ± 0.13
Krippendorff's α	Final vs GPT vs Claude	0.47 ± 0.14	0.48 ± 0.14	0.69 ± 0.16	0.65 ± 0.16	0.39 ± 0.12	0.23 ± 0.22	0.41 ± 0.09
% Agreement	Final vs Claude	0.85 ± 0.06	0.86 ± 0.08	0.97 ± 0.02	0.96 ± 0.03	0.87 ± 0.08	0.70 ± 0.11	0.84 ± 0.05
% Agreement	Final vs GPT	0.84 ± 0.11	0.82 ± 0.14	0.96 ± 0.04	0.93 ± 0.07	0.85 ± 0.10	0.67 ± 0.13	0.80 ± 0.07
% Agreement	Final vs GPT vs Claude	0.84 ± 0.08	0.83 ± 0.12	0.96 ± 0.03	0.94 ± 0.05	0.87 ± 0.08	0.74 ± 0.11	0.82 ± 0.06

D.1 LLM Hallucination

Because our prompt spans multiple dimensions, the LLM sometimes generated labels outside our codebook (beyond minor formatting errors). When these could be mapped to existing categories, we converted them:

```
Extracting > Analyzing
Planning > Analyzing
Recommending > Advising
Predicting > Forecasting
Group > Organizational
```

However, some labels, such as 'Generating' and 'Simulating' for AI roles, could not be directly mapped. We observed fewer such deviations when coding human and AI influence, since each dimension used a separate prompt, though this increased API costs. LLMs also tend to pick up more nuances. For example, if "is developed" is mentioned, LLMs will code for "developers" even if developers are not explicitly discussed in the paper. Unlike human coders, who can receive clarifying instructions mid-task, LLMs require validation before any prompt can be refined. So we keep the initial results but prepare more instructions for human coders.

E Examples of Relevant and Irrelevant Cases

We show examples during our screening process to help readers understand the boundary, highlighting keywords to help them read. Also see Figs. 14 and 15 above.

E.1 Whether to Code for AI and Human Influence

Below is a paper from ACM. It's a user study about trust in AI-assisted decision-making, perfectly aligning with our goal. So we go for AI and human influence coding.

> Who Should I Trust: AI or Myself? Leveraging Human and AI Correctness Likelihood to Promote Appropriate Trust in AI-Assisted Decision-Making
> In AI-assisted decision-making, it is critical for human decision-makers to know when to trust AI and when to trust themselves. However, prior studies calibrated human trust only based on AI confidence indicating AI's correctness likelihood (CL) but ignored humans' CL, hindering optimal team decision-making. To mitigate this gap, we proposed to promote humans' appropriate trust based on the CL of both sides at a task-instance level. We first modeled humans' CL by approximating their decision-making models and computing their potential performance in similar instances. We demonstrated the feasibility and effectiveness of our model via two preliminary studies. Then, we proposed three CL exploitation strategies to calibrate users' trust explicitly/implicitly in the AI-assisted decision-making process. Results from a between-subjects experiment (N=293) showed that our CL exploitation strategies promoted more appropriate human trust in AI, compared with only using AI confidence. We further provided practical implications for more human-compatible AI-assisted decision-making.
> doi: 10.1145/3544548.3581058

Below is a paper from NeurIPS making theoretical contributions. It's highly relevant, but the experiment is based on existing datasets, not a user study. We couldn't code for AI and human influence. The decision type is also very abstract.

> Human-aligned calibration for AI-assisted decision making
> Whenever a binary classifier is used to provide decision support, it typically provides both a label prediction and a confidence value. Then, the decision maker is supposed to use the confidence value to calibrate how much to trust the prediction. In this context, it has been often argued that the confidence value should correspond to a well calibrated estimate of the probability that the predicted label matches the ground truth label. However, multiple lines of empirical evidence suggest that decision makers have difficulties at developing a good sense on when to trust a prediction using these confidence values. In this paper, our goal is first to understand why and then investigate how to construct more useful confidence values. We first argue that, for a broad class of utility functions, there exists data distributions for which a rational decision maker is, in general, unlikely to discover the optimal decision policy using the above confidence values—an optimal decision maker would need to sometimes place more (less) trust on predictions with lower (higher) confidence values. However, we then show that, if the confidence values satisfy a natural alignment property with respect to the decision maker's confidence on her own predictions, there always exists an optimal decision policy under which the level of trust the decision maker would need to place on predictions is monotone on the confidence values, facilitating its discoverability. Further, we show that multicalibration with respect to the decision maker's confidence on her own prediction is a sufficient condition for alignment. Experiments on a real AI-assisted decision making scenario where a classifier provides decision support to human decision makers validate our theoretical results and suggest that alignment may lead to better decisions.
> link: https://proceedings.neurips.cc/paper_files/paper/2023/hash/2f1d1196426ba84f47d115cac3dcb9d8-Abstract-Conference.html

E.2 Borderline Cases

In this section, we present two examples for boundary cases. Both are borderline, but we include one and exclude the other.

Below is a paper contributing a fairness framework. The paper does not explicitly address decision-making, but "fair classification" is actually algorithmic decision-making. So this is a borderline case we still include in the first screening, and didn't include in rescreening.

> Costs and Benefits of Fair Representation Learning
> Machine learning algorithms are increasingly used to make or support important decisions about people's lives. This has led to interest in the problem of fair classification, which involves learning to make decisions that are non-discriminatory with respect to a sensitive variable such as race or gender. Several methods have been proposed to solve this problem, including fair representation learning, which cleans the input data used by the algorithm to remove information about the sensitive variable. We show that using fair representation learning as an intermediate step in fair classification incurs a cost compared to directly solving the problem, which we refer to as the cost of mistrust. We show that fair representation learning in fact addresses a different problem, which is of interest when the data user is not trusted to access the sensitive variable. We quantify the benefits of fair representation learning, by showing that any subsequent use of the cleaned data will not be too unfair. The benefits we identify result from restricting the decisions of adversarial data users, while the costs are due to applying those same restrictions to other data users.
> doi: https://doi.org/10.1145/3306618.3317964

Below is a paper contributing an algorithm for lane-changing in automated driving. It looks indirect to decision-making but automated driving's lane changing is typically considered an HAID. So this is a borderline case we didn't include in the first screening, but included in rescreening.

> Lane Change Intention Awareness for Assisted and Automated Driving on Highways
> Today the automotive industry faces a robust trend toward assisted and automated driving. The technology to accomplish this ambition has evolved rapidly over the last few years, and yet there are still a lot of algorithmic challenges left to make an automation of the driving task a safe and comfortable experience. One of the main remaining challenges is the comprehension of the current traffic situation and the anticipation of all traffic participants' future driving behavior, which is needed for the technical system to obtain situation awareness: an indispensable foundation for successful decision-making. In this paper, a prediction framework is presented that is able to infer a driver's maneuver intention. This is achieved via a hybrid Bayesian network whose hidden layers represent a driver's lane contentedness. A pre-training of the network's parameters with simulated data provides for human interpretable parameters even after running the expectation maximization algorithm based on data gathered on German highways. Moreover, the future driving path of any traffic participant is predicted by solving an optimal control problem, whereby the parameters of the optimal control formulation are found via inverse reinforcement learning.
> doi: 10.1109/TIV.2019.2904386

E.3 Excluding Reasons

In this section, we list our reasons for excluding a paper. Each reason has an example. In practice, one paper may deserve multiple reasons, but we ensure each excluded paper has at least one documented reason.

Reason 1: not HAID

This paper is about networks between cars. Humans are not involved. So this is not HAID (or no human).

> POMDP-Based Decision Making for Fast Event Handling in VANETs
> Malicious vehicle agents broadcast fake information about traffic events and thereby undermine the benefits of vehicle-to-vehicle communication in vehicular ad-hoc networks (VANETs). Trust management schemes addressing this issue do not focus on effective/fast decision making in reacting to traffic events. We propose a Partially Observable Markov Decision Process (POMDP) based approach to balance the trade-off between information gathering and exploiting actions resulting in faster responses. Our model copes with malicious behavior by maintaining it as part of a small state space, thus is scalable for large VANETs. We also propose an algorithm to learn model parameters in a dynamic behavior setting. Experimental results demonstrate that our model can effectively balance the decision quality and response time while still being robust to sophisticated malicious attacks.
> doi: 10.1609/aaai.v32i1.11577

Reason 2 too indirect to HAID

This paper develops a DL framework that could be used for travelers' decision-making, but the paper doesn't focus on decision-making and therefore the connection is too weak.

> A hybrid deep learning approach for urban expressway travel time prediction considering spatial-temporal features
> abstract: Travel time is an effective measure of roadway traffic conditions which enables travelers to make smart decisions about departure time, route choice and congestion avoidance. Recent years have witnessed numerous successes of deep learning neural networks in the domains of artificial intelligence (AI). Motivated by the dominant performance of convolution neural networks (CNNs) and long short-term memory neural networks (LSTMs), and with consideration of the spatial-temporal features, this study attempts to develop a hybrid deep learning framework fusing CNNs and LSTMs to forecast the travel time on urban expressways. A 2-dimension deep CNNs is exploited to capture spatial features of traffic states, and LSTMs are utilized to excavate the temporal correlation of travel time series. Then, these spatial-temporal features are fed into a linear regression layer. The travel time forecasting is achieved by fusing these abstract traffic features in a hybrid deep learning framework. The proposed approach is investigated on Ring 2, a 33km urban expressway of Beijing, China. The results demonstrate the advantage of the proposed method, as well as its feasibility and effectiveness compared with other prevailing parametric and nonparametric algorithms.
> doi: 10.1109/ITSC.2017.8317889

Reason 3: too speculative *(usually suggesting the work can be used in decision-making)*

This reason is applied when the paper speculates that their work can be used for decision-making (in discussion).

How molecular imaging will enable robotic precision surgery
Molecular imaging is one of the pillars of precision surgery. Its applications range from early diagnostics to therapy planning, execution, and the accurate assessment of outcomes. In particular, molecular imaging solutions are in high demand in minimally invasive surgical strategies, such as the substantially increasing field of robotic surgery. This review aims at connecting the molecular imaging and nuclear medicine community to the rapidly expanding armory of surgical medical devices. Such devices entail technologies ranging from artificial intelligence and computer-aided visualization technologies(software) to innovative molecular imaging modalities and surgical navigation(hardware). We discuss technologies based on their role at different steps of the surgical workflow, i. e. , from surgical decision and planning, over to target localization and excision guidance, all the way to(back table) surgical verification. This provides a glimpse of how innovations from the technology fields can realize an exciting future for the molecular imaging and surgery communities.
doi: 10.1007/s00259-021-05445-6

Also, this is a review paper.

Reason 4: no human

This reason could be applied when the paper uses decision as a technical term, and the involvement of humans is unclear.

QINet: Decision Surface Learning and Adversarial
Enhancement for Quasi-Immune Completion of Diverse Corrupted Point Clouds
In point cloud completion task, most previous works fail to deal with diverse corrupted point clouds with large missing areas. Meanwhile, they are restricted by discrete point clouds lacking smooth surfaces to represent an object, and the resolution of generated point clouds is fixed once their networks are determined. In addition, the evaluation metrics are not specific for this task. Thus, we propose an innovative quasi-immune completion architecture of point cloud called QINet in this article, which is inspired by the artificial immunization process in biology. Specifically, to increase robustness and adaptation of the model, we conceive a mask algorithm named onion-peeling (OP) to generate diverse corrupted inputs. Meanwhile, two proposed modules are combined together to produce flexible resolution of point clouds, namely, the decision surface learning and adversarial enhancement for the latent representation recovery. The first module transforms point clouds to surfaces with a continuous decision boundary function, while the second module is applied to deduce complete surface from corrupted point cloud by the cooperation of reinforcement learning (RL) and latent generative adversarial network (GAN). Besides, we evaluate the shortcomings of the existing methods and present two novel metrics to support multifaceted comparisons. Experimental results verify that our approach can generate continuous 3-D shapes with optional resolutions compared with other approaches, and achieves competitive results both quantitatively and qualitatively.
doi: 10.1016/j.trc.2023.104240

Reason 5: not human decision

This reason is similar to Reason 4, usually when the term "decision" refers to model decision. This paper discusses algorithmic decisions that are not designed for humans at all.

Offloading and Resource Allocation With General Task Graph in Mobile Edge Computing: A Deep Reinforcement Learning Approach
In this paper, we consider a mobile-edge computing (MEC) system, where an access point (AP) assists a mobile device (MD) to execute an application consisting of multiple tasks following a general task call graph. The objective is to jointly determine the offloading decision of each task and the resource allocation (e.g., CPU computing power) under time-varying wireless fading channels and stochastic edge computing capability, so that the energy-time cost (ETC) of the MD is minimized. Solving the problem is particularly hard due to the combinatorial offloading decisions and the strong coupling among task executions under the general dependency model. Conventional numerical optimization methods are inefficient to solve such a problem, especially when the problem size is large. To address the issue, we propose a deep reinforcement learning (DRL) framework based on the actor-critic learning structure. In particular, the actor network utilizes a DNN to learn the optimal mapping from the input states (i.e., wireless channel gains and edge CPU frequency) to the binary offloading decision of each task. Meanwhile, by analyzing the structure of the optimal solution, we derive a low-complexity algorithm for the critic network to quickly evaluate the ETC performance of the offloading decisions output by the actor network. With the low-complexity critic network, we can quickly select the best offloading action and subsequently store the state-action pair in an experience replay memory as the training dataset to continuously improve the action generation DNN. To further reduce the complexity, we show that the optimal offloading decision exhibits an one-climb structure, which can be utilized to significantly reduce the search space of action generation. Numerical results show that for various types of task graphs, the proposed algorithm achieves up to 99.1% of the optimal performance while significantly reducing the computational complexity compared to the existing optimization me...
doi: 10.1109/TWC.2020.2993071

Reason 5: no AI

This reason could be applied when the AI terms is actually referring to a human cognitive process. Common in APA (psychology papers).

Behavioral and neurobiological mechanisms of punishment: implications for psychiatric disorders
Punishment involves learning about the relationship between behavior and its adverse consequences. Punishment is fundamental to reinforcement learning, decision-making and choice, and is disrupted in psychiatric disorders such as addiction, depression, and psychopathy. However, little is known about the brain mechanisms of punishment and much of what is known is derived from study of superficially similar, but fundamentally distinct, forms of aversive learning such as fear conditioning and avoidance learning. Here we outline the unique conditions that support punishment, the contents of its learning, and its behavioral consequences. We consider evidence implicating GABA and monoamine neurotransmitter systems, as well as corticostriatal, amygdala, and dopamine circuits in punishment. We show how maladaptive punishment processes are implicated in addictions, impulse control disorders, psychopathy, anxiety, and depression and argue that a better understanding of the cellular, circuit, and cognitive mechanisms of punishment will make important contributions to next generation therapeutic approaches.
doi: 10.1038/s41386-018-0047-3

Reason 6: can't see path from HAI to decision-making

This paper predicts human decisions (HAI) but we really can't see how this is connected to either humans making decisions or AI replacing human decisions.

> Conversational Decision-Making Model for Predicting the King's Decision in the Annals of the Joseon Dynasty
> Styles of leaders when they make decisions in groups vary, and the different styles affect the performance of the group. To understand the key words and speakers associated with decisions, we initially formalize the problem as one of predicting leaders' decisions from discussion with group members. As a dataset, we introduce conversational meeting records from a historical corpus, and develop a hierarchical RNN structure with attention and pre-trained speaker embedding in the form of a, Conversational Decision Making Model (CDMM). The CDMM outperforms other baselines to predict leaders' final decisions from the data. We explain why CDMM works better than other methods by showing the key words and speakers discovered from the attentions as evidence.
> doi: 10.18653/v1/D18-1115

Reason 7: not focus on decision-making

This reason applies when decision-making is just one of the goals. The example below is a review paper mentioning clinical decision-making, but its main focus is on the imaging method. Also, it's a review paper.

> Multiparametric cardiovascular magnetic resonance approach in diagnosing, monitoring, and prognostication of myocarditis
> Myocarditis represents the entity of an inflamed myocardium and is a diagnostic challenge caused by its heterogeneous presentation. Contemporary noninvasive evaluation of patients with clinically suspected myocarditis using cardiac magnetic resonance (CMR) includes dimensions and function of the heart chambers, conventional T2-weighted imaging, late gadolinium enhancement, novel T1 and T2 mapping, and extracellular volume fraction calculation. CMR feature-tracking, texture analysis, and artificial intelligence emerge as potential modern techniques to further improve diagnosis and prognostication in this clinical setting. This review describes the evidence surrounding different CMR methods and image postprocessing methods and highlights their values for clinical decision making, monitoring, and risk stratification across stages of this condition.
> doi: 10.1016/j.jcmg.2021.11.017

Reason 8: no connection between human and AI

This paper proposes to use AI to analyze gestures, posture, facial expressions, and vocal expressions. The AI is not directly connected to humans, or at least such a connection is too weak.

> The promise of social signal processing for research on decision-making in entrepreneurial contexts
> In this conceptual paper, we demonstrate how modern data science techniques can advance our understanding of important decisions in the context of entrepreneurship that involve social interactions. We know that individuals' decision-making is strongly affected by nonverbal behavior. The emerging domain of social signal processing aims at accurate computerized analysis of such behavior. Behavioral cues stemming from, for example, gestures, posture, facial expressions, and vocal expressions can now be detected and analyzed by state-of-the-art technologies utilizing artificial intelligence. This paper discusses and illustrates their potential value for future research on decision-making by entrepreneurs as well as by others yet directly affecting them (e.g., investors). In brief, social signal processing is more accurate and more efficient than conventional research methods and may reveal important characteristics that so far have been omitted in explaining decisions that are vital for firm survival and growth. We derive a total of five propositions from our newly developed conceptual framework, which we hope will be subject to extensive empirical scrutiny in future research.
> doi: 10.2307/48734899

Reason 9: not learning about HAID

This reason often overlaps with others. The paper below, for example, is primarily an algorithmic contribution. While other reasons like Reason 7 could also apply, the core issue is that it offers no new knowledge about HAID.

> Probabilistic verification of fairness properties via concentration
> As machine learning systems are increasingly used to make real world legal and financial decisions, it is of paramount importance that we develop algorithms to verify that these systems do not discriminate against minorities. We design a scalable algorithm for verifying fairness specifications. Our algorithm obtains strong correctness guarantees based on adaptive concentration inequalities; such inequalities enable our algorithm to adaptively take samples until it has enough data to make a decision. We implement our algorithm in a tool called VeriFair, and show that it scales to large machine learning models, including a deep recurrent neural network that is more than five orders of magnitude larger than the largest previously-verified neural network. While our technique only gives probabilistic guarantees due to the use of random samples, we show that we can choose the probability of error to be extremely small. doi: 10.1145/3360544

Reason 10: not talking about human decisions *(just talking about AI decisions)*

This reason is similar to **Reason 5: not human decision**. The decision is purely a model decision.

Repairing Decision-Making Programs Under Uncertainty
The world is uncertain. Programs can be wrong. We address the problem of repairing a program under uncertainty, where program inputs are drawn from a probability distribution. The goal of the repair is to construct a new program that satisfies a probabilistic Boolean expression. Our work focuses on loop-free decision-making programs, e. g. , classifiers, that return a Booleanor finite-valued result. Specifically, we propose distribution-guided inductive synthesis, a novel program repair technique that iteratively(i) samples a finite set of inputs from a probability distribution defining the precondition, (ii) synthesizes a minimal repair to the program over the sampled inputs using an smt -based encoding, and(iii) verifies that the resulting program is correct and is semantically close to the original program. We formalize our algorithm and prove its correctness by rooting it in computational learning theory. For evaluation, we focus on repairing machine learning classifiers with the goal of making them unbiased(fair). Our implementation and evaluation demonstrate our approach's ability to repair a range of programs.
doi: 10.1007/978-3-319-63387-9_9

Reason 11: no clear connection to HAID
This paper is about HAI, but the decision is really a research decision.

Assessing Human-AI Interaction Early through Factorial Surveys: A Study on the Guidelines for Human-AI Interaction
This work contributes a research protocol for evaluating human-AI interaction in the context of specific AI products. The research protocol enables UX and HCI researchers to assess different human-AI interaction solutions and validate design decisions before investing in engineering. We present a detailed account of the research protocol and demonstrate its use by employing it to study an existing set of human-AI interaction guidelines. We used factorial surveys with a 2 × 2 mixed design to compare user perceptions when a guideline is applied versus violated, under conditions of optimal versus sub-optimal AI performance. The results provided both qualitative and quantitative insights into the UX impact of each guideline. These insights can support creators of user-facing AI systems in their nuanced prioritization and application of the guidelines.
doi: 10.1145/3511605

Reason 12: human decisions being studied *(using AI methods to model human decisions)*
In this paper, human irrational decision-making is modeled using Partially Observable Markov Decision Processes. But there is no new knowledge about HAID.

Apparently Irrational Choice as Optimal Sequential Decision Making
In this paper, we propose a normative approach to modeling apparently human irrational decision making(cognitive biases) that makes use of inherently rational computational mechanisms. We view preferential choice tasks as sequential decision making problems and formulate them as Partially Observable Markov Decision Processes(POMDPs). The resulting sequential decision model learns what information to gather about which options, whether to calculate option values or make comparisons between options and when to make a choice. We apply the model to choice problems where context is known to influence human choice, an effect that has been taken as evidence that human cognition is irrational. Our results show that the new model approximates a bounded optimal cognitive policy and makes quantitative predictions that correspond well to evidence about human choice. Furthermore, the model uses context to help infer which option has a maximum expected value while taking into account computational cost and cognitive limits. In addition, it predicts when, and explains why, people stop evidence accumulation and make a decision. We argue that the model provides evidence that apparent human irrationalities are emergent consequences of processes that prefer higher value(rational) policies.
doi: 10.1609/aaai.v35i1.16161

Reason 13: too vague
The paper is too vague about decision-making, and we can't infer what decisions were made.

A Generative Benchmark Creation Framework for Detecting Common Data Table Versions
Multiple versions of the same dataset can exist in a data repository (e.g., data warehouses, data lakes, etc.), mainly because of the interactive and collaborative nature of data science. Data creators generally update existing datasets and upload them as new datasets to data repositories without proper documentation. Identifying such versions helps in data management, data governance, and making better decisions using data. However, there is a dearth of benchmarks to develop and evaluate data versioning techniques, which requires a lot of human effort. Thus, this work introduces a novel framework to generate benchmarks for data versioning using Generative AI (specifically Large Language Models). The proposed framework offers properties that existing benchmarks do not have, including proper documentation, version lineage, and complex transformations generated by an LLM. We also share VerLLM-v1, the first version of the benchmark that features these properties, and compare it to existing benchmarks.
doi: 10.1145/3627673.3679157

Reason 14: too abstract
This is similar to **Reason 13: too vague**, but the decision in the paper is too abstract, and we can't infer what decisions were made.

Branch Ranking for Efficient Mixed-Integer Programming via Offline Ranking-Based Policy Learning
Deriving a good variable selection strategy in branch-and-bound is essential for the efficiency of modern mixed-integer programming(MIP) solvers. With MIP branching data collected during the previous solution process, learning to branch methods have recently become superior over heuristics. As branch-and-bound is naturally a sequential decision making task, one should learn to optimize the utility of the whole MIP solving process instead of being myopic on each step. In this work, we formulate learning to branch as an offline reinforcement learning(RL) problem, and propose a long-sighted hybrid search scheme to construct the offline MIP dataset, which values the long-term utilities of branching decisions. During the policy training phase, we deploy a ranking-based reward assignment scheme to distinguish the promising samples from the long-term or short-term view, and train the branching model named Branch Ranking via offline policy learning. Experiments on synthetic MIP benchmarks and real-world tasks demonstrate that Branch Ranking is more efficient and robust, and can better generalize to large scales of MIP instances compared to the widely used heuristics and state-of-the-art learning-based branching models.
doi: 10.1007/978-3-031-26419-1_23

Reason 15: not relevant
AI in this paper is used in a part not relevant to the research.

Dynamic capabilities framework and its transformative contributions
Dynamic capabilities refer to an organization's ability to integrate, build, and reconfigure internal and external competencies to address a rapidly developing environment. [...] This text was initially drafted using artificial intelligence, then reviewed by the author(s) to ensure accuracy.
doi: 10.1057/s41267-024-00758-8

Reason 16: not firsthand knowledge
This reason was applied when the paper uses decision as motivation but the work doesn't contribute knowledge about HAID.

Towards Learning Contrast Kinetics with Multi-condition Latent Diffusion Models
Contrast agents in dynamic contrast enhanced magnetic resonance imaging allow to localize tumors and observe their contrast kinetics, which is essential for cancer characterization and respective treatment decision-making. However, contrast agent administration is not only associated with adverse health risks, but also restricted for patients during pregnancy, and for those with kidney malfunction, or other adverse reactions. With contrast uptake as key biomarker for lesion malignancy, cancer recurrence risk, and treatment response, it becomes pivotal to reduce the dependency on intravenous contrast agent administration. To this end, we propose a multi-conditional latent diffusion model capable of acquisition time-conditioned image synthesis of DCE-MRI temporal sequences. To evaluate medical image synthesis, we additionally propose and validate the Fréchet radiomics distance as an image quality measure based on biomarker variability between synthetic and real imaging data. Our results demonstrate our method's ability to generate realistic multi-sequence fat-saturated breast DCE-MRI and uncover the emerging potential of deep learning based contrast kinetics simulation. We publicly share our accessible codebase at https://github.com/RichardObi/ccnet and provide a user-friendly library for Fréchet radiomics distance calculation at https://pypi.org/project/frd-score.
doi: 10.1007/978-3-031-72086-4_67

Reason 17: not enough evidence
This is similar to **Reason 3: too speculative**. The paper uses decision-making in discussion but doesn't have evidence for applicability.

Human understandable thyroid ultrasound imaging ai report system — a bridge between ai and clinicians Summary Artificial intelligence (AI) enables accurate diagnosis of thyroid cancer; however, the lack of explanation limits its application. In this study, we collected 10,021 ultrasound images from 8,079 patients across four independent institutions to develop and validate a human understandable AI report system named TiNet for thyroid cancer prediction. TiNet can extract thyroid nodule features such as texture, margin, echogenicity, shape, and location using a deep learning method conforming to the clinical diagnosis standard. Moreover, it offers excellent prediction performance (AUC 0.88) and provides quantitative explanations for the predictions. We conducted a reverse cognitive test in which clinicians matched the correct ultrasound images according to TiNet and clinical reports. The results indicated that TiNet reports (87.1% accuracy) were significantly easier to understand than clinical reports (81.6% accuracy; $p < 0.001$). TiNet can serve as a bridge between AI-based diagnosis and clinicians, enhancing human–AI cooperative medical decision-making.
doi: https://doi.org/10.1016/j.isci.2023.106530

Reason 18: too hard to code
We attempt to code the dimensions, but find it's too hard to code. Also, this paper's main focus is AI readiness, not decision-making.

Ready or Not, AI Comes—An Interview Study of Organizational AI Readiness Factors
Artificial intelligence (AI) offers organizations much potential. Considering the manifold application areas, AI's inherent complexity, and new organizational necessities, companies encounter pitfalls when adopting AI. An informed decision regarding an organization's readiness increases the probability of successful AI adoption and is important to successfully leverage AI's business value. Thus, companies need to assess whether their assets, capabilities, and commitment are ready for the individual AI adoption purpose. Research on AI readiness and AI adoption is still in its infancy. Consequently, researchers and practitioners lack guidance on the adoption of AI. The paper presents five categories of AI readiness factors and their illustrative actionable indicators. The AI readiness factors are deduced from an in-depth interview study with 25 AI experts and triangulated with both scientific and practitioner literature. Thus, the paper provides a sound set of organizational AI readiness factors, derives corresponding indicators for AI readiness assessments, and discusses the general implications for AI adoption. This is a first step toward conceptualizing relevant organizational AI readiness factors and guiding purposeful decisions in the entire AI adoption process for both research and practice.
doi: 10.1007/s12599-020-00676-7

Reason 19: 1 page / 2 pages

After retrieving the full text, we determined that the paper was only 1–2 pages in length, which is insufficient for the analysis required by our review.

Reason 20: tutorial

The abstract shows that the paper is a tutorial at a conference (often a review paper too), and doesn't have a substantial research contribution.

> Psychology-informed Recommender Systems Tutorial
> Recommender systems are essential tools to support human decision-making in online information spaces. Many state-of-the-art recommender systems adopt advanced machine learning techniques to model and predict user preferences from behavioral data. While such systems can provide useful and effective recommendations, their algorithmic design commonly neglects underlying psychological mechanisms that shape user preferences and behavior. In this tutorial, we offer a comprehensive review of the state of the art and progress in psychology-informed recommender systems, i.e., recommender systems that incorporate human cognitive processes, personality, and affective cues into recommendation models, along with definitions, strengths and weaknesses. We show how such systems can improve the recommendation process in a user-centric fashion. With this tutorial, we aim to stimulate more ideas and discussion with the audience on core issues of this topic such as the identification of suitable psychological models, availability of datasets, or the suitability of existing performance metrics to evaluate the efficacy of psychology-informed recommender systems. Besides, we present takeaways to recommender systems on how to build psychology-informed recommender systems. Previous versions of this tutorial were presented, among others, at The ACM Web Conference 2022 and the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR) 2022.
> doi: 10.1145/3523227.3547375

Reason 21: Editorial/Commentary/Reply/Letter/Preface/Discussion

Papers in these genres serve primarily to comment on, introduce, or respond to other work rather than present original research. While they may reference HAID, they lack the empirical findings or theoretical development.

> Reporting use of ai in research and scholarly publication—jama network guidance
> Reports on the use of artificial intelligence (AI) and machine learning, including large language models (LLM), in medical research have intensified in the last year. Although machine learning research began 70 years ago with the conceptual development of artificial neural network algorithms, AI research and use in clinical practice and health care are relatively recent advances. Throughout these developments, JAMA has sought to define the broad scope of discovery and innovation in medical applications of AI and to address potential challenges in its implementation. The journal's 2016 publication of a study of deep learning algorithms for the detection of diabetic retinopathy from fundal photographs is a useful example. The study represented a novel tool that could enable large-scale screening for a key vision-threatening disorder across the world. However, the accompanying Editorial highlighted important challenges, spanning the need for broader patient representativeness, the investment necessary in validating the model in the context of its deployment and subsequent implementation, and whether clinicians would entrust decision-making to AI tools. The Editorial also called attention to concerns regarding AI eventually replacing humans in clinical systems. Although published before the recent AI boom, this study and the comments about it augured many of today's promises and concerns regarding AI in clinical research and practice.
> doi: 10.1001/jama.2024.3471

Reason 22: Early Career Track

IJCAI has a track inviting early-career AI researchers to present based on nominations rather than presenting original research.

> Improving group decision-making by artificial intelligence
> We summarize some of our recent work on using AI to improve group decision-making by taking a unified approach from statistics, economics, and computation. We then discuss a few ongoing and future directions.
> link: https://www.ijcai.org/proceedings/2017/741

Reason 23: Survey/Opinion

Pure survey papers, review articles, and opinion pieces synthesize or comment on existing work rather than present original research. Their broad scope and lack of specific methodological details make them unsuitable for systematic coding. However, sometimes a paper can have both survey/opinion contributions along with other original contributions.

F Prompts for LLMs

F.1 Prompt to Get Venue Ranks

```
System prompt:  I give you a publication venue which we find in {database}. You output its rank (top-tier, mid-tier, or low-tier)
, peer-reviewed (yes or no), area, venue, as well as your confidence.
Your output should be in this format: rank: XXXX || peer-reviewed: XXXX || area: XXXX || confidence: 0-1 || venue: its most
commonly-accepted full name || explanation: 10-80 words.

The venue is {conf}.
```

F.2 Prompt When Abstract is Used

```
System prompt: You are acting as an annotator for a literature review and you should output only codes.

## Task Overview
```

```
Analyze this academic paper's title and abstract to classify ALL aspects: application domain, research contribution type,
decision-making factors, AI roles, and human roles. Output the exact format and DO NOT output anything else.

## DECISION DOMAINS (Select ONE)
- **Healthcare / Medicine / Surgery**
  Examples: public health, nursing, childcare, clinical, medicine, healthcare, surgery, therapist
- **Finance / Business / Economy**
  Examples: finance/investment/loan, business/trading/shopping/marketing, blockchain/bitcoin
- **Education / Teaching / Research**
  Examples: schooling, teaching, academia, research, data analysis, assessments, science
- **Law / Policy / Governance**
  Examples: criminal/law/legal/justice; politics/government/democracy, policymaking, campaign
- **Software / Systems / Security**
  Examples: smart home, programming, software engineering, cloud computing, databases, privacy, internet infrastructure
- **Transportation / Mobility / Planning**
  Examples: city/smart city; driving, traffic, logistics, travel, freight, navigation, flight, aviation
- **Manufacturing / Industry / Automation**
  Examples: manufacturing, supply chain, logistics, robotics, automation, resource allocation
- **Environment / Resources / Energy**
  Examples: water, food,  agriculture, fishery, water, food, energy, electricity, environment
- **Defense / Military / Emergency**
  Examples: homeland security, military, disaster prediction, emergency response, surveillance
- **Media / Communication / Entertainment**
  Examples: press/journalism/writing, sport, game/entertainment, social media/social network, tourism
- **Design / Creativity / Architecture**
  Examples: product design, building architecture, user-centered design, creative industries
- **Everyday / Employment / Public Service**
  Examples: hiring, admissions, caregiving, daycare, social good, public sector, management, everyday tasks, renting,
organizational decision-making, image classification related to everyday tasks
- **Generic / Abstract / Domain-agnostic**
  Definition: General purposes or abstract context; general image classification

## CONTRIBUTION TYPES (Select ONE)
- **Algorithmic contributions**
  Definition: new AI/ML algorithms, models, or computational approaches, very specific thing
- **Empirical contributions**
  Definition: experimental findings, user studies, and observational research results
- **Methodological contributions**
  Definition: new research methods, evaluation frameworks, analytical approaches, and design guidelines/principles
- **Theoretical contributions**
  Definition: conceptual frameworks, Conceptual models, and theories
- **System/Artifact contributions**
  Definition: implemented systems, tools, interfaces, and design solutions
- **Dataset/Benchmark contributions**
  Definition: new datasets, benchmarks, and evaluation resources
- **Survey contributions**
  Definition: literature reviews and synthesis papers, overview of a field
- **Vision contributions**
  Definition: position papers, future directions, and perspective pieces

## DECISION-MAKING FACTORS (Select 1-3)
- **Outcome Quality**
- **Interpretability**
- **Explainability**
- **Agency**
- **Uncertainty**
- **Risk**
- **User Trust**
- **Reliance**
- **Information Quality**
- **Decision Context**
- **User Motivation**
- **Fairness**
- **Ethics/Value Alignment**
- **Personalization**

## AI ROLES (Select 1-2)
```

- **Analyzing**
  Definition: Processes complex information to give it structure and clarity
  Examples include (pattern/action) recognizer, biomarker identifier, (human behavior/prognostic) modeler, (treatment effect) estimator, risk assessment tool, summarizer, (information/knowledge) extractor, report generator, information/option provider, planner, analyzer
- **Forecasting**
  Definition: Uses current and historical data to forecast what will happen in the future. It answers the question, \"What's next ?\"
  Examples include (counterfactual/risk) predictor, forecaster, mortality prediction model, (image) classifier, (skin cancer) detector
- **Explaining**
  Definition: Translates opaque processes into human-readable justifications. \"Why did this happen?\"
  Examples include explainer, reasoner
- **Advising**
  Definition: Actively provides recommendations or guidance for the human to consider;\n
  Examples include recommender, AI teacher, (suggestion/portfolio) generator, (robo-) advisor, tutor, coach, guide, curriculum designer, advice interpreter, decision-making model/algo/system
- **Collaborating**
  Definition: Works alongside a human as a partner in a dynamic, interactive process;
  Examples include collaborator, AI partner, teammate, human-agent team member, hybrid partner, complementary learner, co-pilot, interactive agent, shared autonomy system, assistant/supplement, collaboration aid, conversational agent (CA), chatbot, virtual assistant, dialogue manager, health chatbots
- **Executing**
  Definition: Acts independently to perform a task or make a decision delegated by a human;
  Examples include task performer/delegatee, assigner/matcher, actor, agent, (workflow) optimizer, allocator, scheduler, controller, task automator, resource generator, automated (trader/negotiator), (game/autonomous) player, strategic agent, policy generator, decision-maker, (policy) learner
- **Monitoring**
  Definition: Monitors a system's real-time performance, status, and safety;
  Examples include (trajectory/policy) evaluator, supervisor, overseer, (quality/reliability) assessor, performance monitor, system health diagnostician
- **Auditing**
  Definition: Checks and validates a system against established rules, ethics, and standards;
  Examples include (fairness/policy) auditor, ethical verifier, (bias/discrimination) detector, formal modeler

HUMAN ROLES (Select 1-2)
- **Decision-maker**
  Definition: Individuals who use or collaborate with AI to make (final) decisions or take actions;
  Examples: decision-maker, investigator, user, collaborator, learner, delegator
- **Decision-subject**
   Definition: Individuals are directly/personally affected/impacted by the decisions made by humans/AI;
   Examples: (cancer) patients, (loan/job) applicants, defendant / inmates, candidate / students, customer / consumer, decision subjects(defendants, citizens, authors, ...)
- **Developer**
  Definition: Technical professionals who design, build, fine-tune, deploy, or maintain the AI system;
  Examples: AI (developers/software engineers), data scientist/engineer, ML model engineers/practitioners, (system) designers, database administrators, technical experts, researcher
- **Guardian**
  Definition: Individuals and institutions that set the rules, policies, and ethical guardrails for AI systems and govern, supervise, regulate, audit, or enforce AI accountability & ethics
  Examples: regulator, policy-maker, government, (independent) auditor, AI ethicist, moral philosophers, research ethics committees, lawyers, judges, courts, parole board, governing body, supervisor, evaluator, analyst/auditor
  - **Knowledge provider**
  Definition: People who provide the specialized data, expertise, or judgment used to train and validate the AI;
  Examples: (domain/subject matter/medical/legal) experts, annotator, labeler, manual coder, survey participant, crowd workers, human subjects, human as inspiration, human as benchmark / guide, teacher
- **Stakeholder**
  Definition:The broader/larger community or social groups indirectly impacted by the AI's deployment and societal consequences
  Examples: community, citizens, public, families, employees, company stakeholders, (vulnerable/protected) groups, society

Output Format
COMPLETE CLASSIFICATION:
Decision_Domain: [domain]
Contribution_Type: [type]
Decision_Factors: [selected facor(s)]
AI_Roles: [selected role(s)]
Human_Roles: [selected role(s)]

```
**Paper to Analyze:**
Title: {title}
Abstract: {abstract}
```

F.3 Prompt When Full-text is Used

```
System prompt: You are acting as an annotator for a literature review analyzing human-AI decision-making. Output only the
requested codes and descriptions.

## Task Overview
Analyze this academic paper to classify the main DECISION task being studied. Output ONLY the requested format.
## DECISION TYPES (Select ONE unless there are multiple decision tasks; when you can't find a decision, output NA)
- **Individual**
    Definition: Decisions made by a single person for themselves
    Examples: personal diabetes management, purchasing products, accepting AI shopping advice, lane-changing, selecting fashion
items/movies/university, travel planning, farmers' planting decisions, program library selection, creative support tools
- **Operational**
    Definition: Decisions embedded in day-to-day routines and practices and made for others
    Examples: clinical diagnosis, coordinating connected fleets, loan approval, hiring/firing, admissions, resource allocation,
policy development, grading students, identifying harmful posts, production scheduling, load shedding, code review, military
operators
- **Organizational**
    Definition: Decisions made within or on behalf of an organization (e.g. school, company, group)
    Examples: emergency department management, COVID vaccine allocation, supply chain management, aviation operations, workforce
management, student recruitment, game management, managing suppliers/retailers, emergency facility planning
- **Institutional**
    Definition: Decisions that establish standards or norms, usually across organizations
    Examples: governance frameworks, legitimacy of algorithmic decisions in markets/public services/schools, social media
filtering, digital twins, climate forecasts, developing defensive strategies under attack, air defense planning
## Output Format
DECISION_TYPE: [selected type, separated by comma]

## Task Overview
Analyze this academic paper to identify HOW AI INFLUENCES HUMANS in decision-making. Focus on the human-subjects quantitative/
qualitative experiments, or system design. Output ONLY the requested format.
## AI INFLUENCE CATEGORIES (Select 0-3 that apply; when you can't find AI influence, output NA)
- **Alter decision outcomes**
    Definition: AI changes the final decisions humans make
    Examples: improving accuracy, improving speed/response rate, reducing speed, achieving complementary human-AI performance,
changing decisions, following AI recommendations, improving performance, increasing decision quality
- **Change trust**
    Definition: AI affects human trust, reliance, or faith in the system
    Examples: trust (except moral trust), reliance, aversion, faith in AI's capabilities, calibrating trust, over-reliance, under
-reliance
- **Change cognitive demands**
    Definition: AI alters the mental effort or cognitive resources required
    Examples: cognitive load, human effort, burden, recall, retention, decision time, skills, mental workload
- **Change affective-perceptual**
    Definition: AI influences feelings, confidence, or subjective experiences
    Examples: confidence, feelings, satisfaction, emotional states, perceived control
- **Restrict human agency**
    Definition: AI limits human autonomy or control in decision-making
    Examples: autonomy, agency, control, freedom to choose, self-determination
- **Shift responsibility**
    Definition: AI changes perceptions of who is accountable for decisions
    Examples: responsibility, blame, accountability, moral trust, liability attribution
- **Shape ethical norms**
    Definition: AI influences fairness perceptions, biases, or ethical standards
    Examples: reducing bias, ensuring fairness, detecting direct/indirect biases, increasing agreement with model biases,
increasing acceptance of unethical advice, reinforcing perceived fairness
## Output Format
AI_INFLUENCE: [selected categories, separated by comma]

Analyze this academic paper to identify HOW HUMANS INFLUENCE AI in decision-making. Focus on the human-subjects quantitative/
qualitative experiments, or system design. Output ONLY the requested format.
```

```
## HUMAN INFLUENCE CATEGORIES (Select 0-3 that apply; when you can't find AI influence, output NA)
- **Update AI competence**
    Definition: Humans influence system robustness and perceived competence
    Examples: improving performance, improving accuracy, enhancing interpretability, returning predictable outputs, improving AI
performance, keeping systems transparent, setting preferences/priorities, model training, fine-tuning
- **Change AI responses through feedback**
    Definition: Direct ways humans refine and shape AI outputs
    Examples: refining prompts, learning from feedback, providing desired information to trust AI, human-adjusted refinements,
providing corrections, providing clarifications, increasing depth of information, iterative refinement
- **Shape AI for accountability**
    Definition: Humans establish governance, oversight, or ethical constraints on AI
    Examples: setting ethical guidelines, establishing governance frameworks, defining fairness constraints, implementing
oversight mechanisms, auditing AI decisions, setting policy constraints
## Output Format
HUMAN_INFLUENCE: [selected categories, separated by comma]

**Paper Content:** {text extracted from pdf (first 80 pages)}
```

G Venues Included in Searching

Here we list which venues we include if the conference has multiple tracks.

AAAI

AAAI main conference, IAAI, EAAI, AIIDE, HCOMP, ICAPS, ICWSM, KR (https://proceedings.kr.org), SoCS, Fall & Spring Symposia, Flairs

ACL Anthology/PMLR/MLR: We started with all venues, but the venue rank checking will narrow down to the top-tier ones.

NeurIPS: 2024: Competition_Track, Datasets_and_Benchmarks_Track, Conference; 2023: Datasets_and_Benchmarks_Track, Conference

H List of Supplementary Materials

Our corpus, source code, annotation tool, and codebook are at https://doi.org/10.17605/OSF.IO/U5QJH. Below we show a list of content. The interactive website is at https://fig-x.github.io/haid/.

- `appendices.pdf`: machine learning model specifications and performance metrics, feature analysis and robustness assessments, LLM-assisted coding methodology and inter-rater reliability results, examples for inclusion/exclusion boundaries, complete prompts used for automated coding, etc.

Folder: Data Collection

The following scripts and documents were used in the data collection stage.

- `disgareement_alignment.xlsx`: This document contains two sheets:
  Disagreement: venue-rank votes where the two LLMs disagree (but at least one votes top-tier).
  Alignment: comparison with human judgments.
- `IEEE/`, `Neurips/`, `Science/`, `Wiley/`: Representative web scraping scripts. Other scripts follow similar structures.

Folder: Analysis

- `1-clean_up_databases.[Rmd|html]`: Script for Quality Assessing (Iteration 1). Also calls GPT and Claude to judge venue ranks.
- `2-combine_databases.[Rmd|html]`: Script for Quality Assessing (Iteration 2), narrowing down to top-tier venues.
- `3-label-prediction.ipynb`, `3-label-prediction-2class (early exploration).ipynb`, `3-label-prediction (Colab; early exploration).pdf`: Five short text–classification models explored for Abstract Screening (Iteration 2): SVM, Random Forest, Logistic Regression, RNN, and BERT. This also extracts feature coefficients from the SVM.
- `4 ... 8`: Internal scripts used to check progress.
- `9-get-llm-coding_abstract.Rmd`, `9-get-llm-coding_full.Rmd`: Scripts for generating LLM coding results.
- `10-cluster-analysis.[Rmd|html]`: Cluster analysis on all abstracts; generates embeddings.
- `11-error-analysis.[Rmd|html]`: Error analysis, feature analysis, and bootstrapping to estimate the effects of miss-outs.
- `12-all-model.py`: We trained again on our historical partial screening results to get Fig. 7.
- `13-figures-and-distritbuion.[Rmd|html]`: Distribution of codes and trend analysis.
- `14-IRR.[Rmd|html]`: Compute inter-rater reliability and generate figures.
- `AID.Rproj`: R project configuration.
- `all_papers_raw.csv`: The 39K papers after filtering out venues.
- `theme.R` & `tool.R`: Helper files.
- Subfolder: vis/ `fig-2.html`: generates Figure 2 (overview of the process)
- Subfolder: output_dataset/ datasets for performance trends and feature analysis used in `11-error-analysis.[Rmd|html]`

Folder: Corpus and Overview

- `Codebook.pdf`: Final mapping and definitions of codes.
- `final_paper_dr.csv`: Embeddings and clusters for abstracts.
- `final_papers.csv`: Codes, metadata, titles, abstracts, and authors for all papers.
- `Overview.html`: Interactive HTML for the corpus, rendered from the two `.csv` files. See https://fig-x.github.io/haid/.

.zip: Annotation tool

The `paper-annotation-tool` provides a lightweight end-to-end system for managing paper annotation through a browser interface backed by Google Apps Script.

- Backend. A script (`google-apps-script/Code.gs`) exposes API endpoints for reading and writing annotations to Google Sheets.
- Frontend. A React-based interface located in `src/` includes reusable UI components, annotation schema and API settings, etc.
- Public assets. The HTML shell (`public/index.html`) in which the React app is rendered.
- Documentation. Setup and deployment guides (`README.md`, `QUICKSTART.md`, `DEPLOYMENT.md`, `SHEETS_SETUP.md`).
- Project configuration. Node package metadata and development settings (`package.json`, `.gitignore`).