ISSN: 2473-4901
v12 n6957

Teaching Case

Auditing AI: Using Excel to Detect Bias in a Hiring System

Setareh Seraj
setareh.seraj@baruch.cuny.edu

Farzam Lahabi
farzam.lahabi@baruch.cuny.edu

Baruch College (CUNY)
New York, NY

Hook

Can you rely on an AI hiring tool to be fair? This case challenges students to audit an AI algorithm's job-screening decisions with Excel to uncover an unintentional bias and then analyze the consequences for organizations by using ethical frameworks.

Abstract

Organizations increasingly use artificial intelligence (AI) tools to screen job applicants, and these algorithms have the potential to replicate biases. This teaching case helps with data-analysis skills and ethical reasoning by having students audit an AI resume-screening tool used at a fictional company, TechNova Solutions, following a discrimination complaint. Students test the fairness of the AI tool toward applicants by applying basic Microsoft Excel techniques to a provided dataset of 200 applicants. This exercise is a simulated algorithmic audit rather than an audit of a real case study, so the data is synthetic and the bias is reported to students in advance. The analysis follows the disparate-impact standard in employment law, which is operationalized with the four-fifths rule. Students then interpret their findings through the four ethical frameworks. Therefore, the case consists of a quantitative audit and a defensible ethical judgment about the findings. On a human-graded skills test, students find that men and women are equally skilled, but the AI gives women lower scores, resulting in a disparity in interview rates that falls below the four-fifths threshold. After that, they think about what the organization should do ethically. This case is designed for introductory information systems and business analytics courses where business ethics is typically included as well, and it does not require programming. This case prepares students for AI-augmented professions and helps them identify and evaluate algorithmic bias.

Keywords: AI bias, algorithmic fairness, business ethics, data literacy, Excel, hiring analytics

Teaching Case
Auditing AI: Using Excel to Detect Bias in a Hiring System

Setareh Seraj and Farzam Lahabi

1. INTRODUCTION

Recruiting, one of the most important decisions a company makes, is now subject to data-driven decision-making (Raghavan et al., 2020). Employers prefer automated methods to screen or rank applicants to save time and cut costs, as well as to land on a more impartial and consistent decision than human recruiters offer (Köchling & Wehner, 2020; Köchling et al., 2021). The intuition is that a machine is more impartial than a human, as it has never seen a candidate's name or face and cannot discriminate (Lepri et al., 2018, as cited in Köchling & Wehner, 2020).

However, this intuition has not held. Because AI screening algorithms learn from past hiring data, they absorb patterns of past human judgments, including discriminatory ones, and then apply them at scale (Njoto et al., 2025; Chen, 2023). After discovering that an experimental recruiting tool penalized resumes with the word "women's" and devalued graduates of two women's universities, a major technology business decided to abandon it (Dastin, 2018). The bias resulted from the data and the method; no coder created a rule that favored men. For students studying information systems, this is the deeper lesson: a system can cause harm to actual people even if it was not intended to, and identifying that harm using common tools is a fundamental data-literacy ability (Barocas & Selbst, 2016).

A worker is less likely to be replaced by AI itself than by a worker who knows how to use AI well (World Economic Forum, 2023). Recognizing AI system failures and biases is important knowledge that comes from practice rather than from lectures. However, most introductory courses treat algorithmic bias as a topic of discussion, so students learn to define the problem without learning to detect it.

This teaching case fills that gap. By using Microsoft Excel and a dataset of 200 job applicants, students can act as analysts investigating the fairness of an AI hiring tool and then reflect on what the organization should do ethically. To do so, they need the technical ability to identify a discrepancy in the data by applying the four-fifths rule used in U.S. employment discrimination law, and the ethical judgment to determine what the organization owes the individuals impacted. In order to help students transition from what happened to what should be done, the case combines a practical data audit with an organized ethical analysis drawn from four traditions: utilitarian, deontological, justice-based, and virtue-based. In this teaching case, students ran a practice audit on the bias-built and synthetic data to learn the method, so the goal is not to reproduce the full difficulty of a real investigation.

2. BACKGROUND

2.1 Why AI Systems Discriminate in Hiring

Employing algorithmic decision-making in hiring offers many benefits for organizations, such as reducing costs and time, boosting productivity, and mitigating human biases that may skew selection, thereby, in principle, increasing both the efficiency and the fairness of the process (Köchling & Wehner, 2020; Köchling et al., 2021). Discrimination is the unfair treatment of groups based on attributes such as gender, age, or ethnicity, rather than on qualitative distinctions such as individual performance (Arrow, 1973, as cited in Köchling & Wehner, 2020). The primary cause of biased recruiting algorithms is the training data. When algorithms are trained on erroneous (Kim, 2017), biased (Barocas & Selbst, 2016), or unrepresentative input data, they will duplicate biased decisions. Predictive hiring systems carry the inequality and biases of their training datasets, such as historical hiring decisions (Njoto et al., 2025). Chen (2023) locates the source of algorithmic bias in limited raw datasets and in the assumptions of the designers who build the models. According to this view, algorithmic bias is the set of recurring, systematic errors that produce unfair outcomes based on legally protected traits such as gender and

race (Jackson, 2021, as cited in Chen, 2023). These systems may even magnify the biases in their inputs rather than just replicating them (Lloyd, 2018, as cited in Chen, 2023).

Crucially, a model can be very accurate while still discriminating. By examining 10,000 video clips, Köchling et al. (2021) discovered that because gender and ethnicity were underrepresented in the training data, the algorithms overestimated or underestimated the likelihood of inviting members of those groups to an interview, reproducing preexisting disparities. The design of an algorithm can therefore make it systematically biased against women in male-dominated fields, particularly when the training data or the very definition of a "qualified" candidate already favors men (Ip, 2025).

To address this issue, employment law uses a disparate-impact criterion to recognize such results. The U.S. Equal Employment Opportunity Commission's four-fifths rule considers a selection rate below 80% of the highest group's rate as potential evidence of adverse impact (U.S. Equal Employment Opportunity Commission [EEOC], 1978).

2.2 Four Ethical Frameworks

An audit can discover that an AI system might produce disparities, but it cannot determine what the organization owes the applicants affected by these disparities, and the four ethical traditions used in this case answer that question differently. Utilitarianism looks to consequences and asks us which course of action yields the greatest benefits for the greatest number of people (Gustafson, 2013). This means that, from welfare-economic perspectives, the optimal option is the one that maximizes the combined weighted utilities of individuals (Harsanyi, 1955). However, deontological ethics shifts attention to the action rather than its consequences. According to Kant's Formula of Universal Law, an action is immoral if the principle behind it cannot be universalized (Scharding, 2015). For example, a screening rule is not ethical even when the consequence of hiring improves the organization's profits. From the justice perspective, distribution deserves attention, and the focus is on whether all applicants are treated similarly and whether opportunity is allocated fairly across all groups. For this matter, Rawls (1971) offers the veil of ignorance as a lens for asking whether a screening rule would be acceptable if all applicants were screened without knowing which group they belong to. The last ethical framework, virtue ethics, moves the focus from the action to the person performing it. It asks what someone with good character would do, and it treats virtue as a stable property of character rather than a set of observable behaviors (Bright et al., 2014; Sison & Ferrero, 2015).

2.3 Related Teaching Cases and Activities

Information systems educators are adding teaching cases and activities about AI. For instance, Sharp et al. (2026) classify eighteen peer-reviewed ISCAP publications from 2022 to 2025 that address AI in teaching and learning, using Zuboff's automate, informate, and transformate categories.Most fell into the informate category, where the tool supports student comprehension and skill development. Fewer fell into the automate or transformate categories. In one case, students write prompts, produce output, and learn what a generative tool can do for them (Firth & Triche, 2024). In the second case, students use an AI tool to critique and improve dashboards they built themselves (Ariyachandra, 2026). In the last case, students evaluate ChatGPT's reading of a visualization that subtly violates a linear regression assumption and judge whether the AI detected it (Larson et al., 2026). This case extends that approach in two ways. First, the system under review is a deployed decision tool rather than a single AI response; second, the question is not whether the AI is correct but whether it is fair to the people it affects.

What is missing is a case that asks students to identify AI biases and analyze them through ethical frameworks. It is not yet common practice to ask students to test a deployed AI model's output against a human benchmark. The activities that build that skill usually require programming or statistics that introductory students lack, and the activities introductory students can complete stop at discussion.

3. DESIGN

This case is designed to move students from being able to define algorithmic bias to being able to find it. Introductory courses cover the idea that a system can produce unfair outcomes and show students the processes involved, so they are able to point to the problem. Few of these cases have ever opened

a dataset and shown students what has happened. The case presented here walks students through a complete audit, beginning with establishing that a gap in outcomes exists. Our case helps students identify how AI produces disparities and explain how these disparities affect what the organization owes the applicants. On completing this case, students will be able to:
1. Compute and compare group outcomes in Excel using pivot tables and apply the four-fifths rule to highlight adverse impact.
2. Distinguish a difference in outcomes from a difference in qualifications by comparing groups on multiple independent measures.
3. Use the CORREL function to identify what drives an outcome and to detect a variable's hidden relationship to a protected characteristic.
4. Apply VLOOKUP, filtering, charting, and conditional formatting to investigate and communicate a data question.
5. Explain how an AI system can produce biased results even when no bias was intended and the protected attribute was withheld.
6. Analyze an organizational decision through utilitarian, deontological, justice, and virtue frameworks and defend a recommendation that names its trade-offs.

The case design is mainly shaped by two constraints. The first is the proficiency level of undergraduate students in introductory information systems or business analytics courses regarding coding, ethics, and basic statistics. Therefore, the entire audit had to run on Excel tools that an introductory course already teaches: pivot tables, a few simple functions, one correlation, VLOOKUP, charts, and conditional formatting. Although some exposure to correlation helps, it can be introduced during the exercise itself. To prepare students for the ethical part of the case, the four ethical frameworks are defined inside it and can be taught in a single class period. Students do not need any prior familiarity with AI tools because they will be auditing a tool rather than using one. The second constraint is that the case's final goal is to make a top-level organizational judgment about the complaint rather than simply report the findings. The disparity reported by students is a statistical problem; what makes it an information systems problem is that someone must weigh the tool's efficiency against the harm it caused, which no amount of analysis can do on its own. As a result, the case needs an organization with something at stake, a person who was harmed and complained, and a role in which the student is responsible for recommending rather than merely observing.

To address these constraints, the case is built around a fictional firm. TechNova Solutions is a mid-sized technology-services company with roughly 900 employees. Facing a surge in job applications, the VP of Human Resources approved the purchase and use of an AI resume-screening platform, TalentSift. This platform scores each applicant from 0 to 100. TechNova uses a 70 cutoff to advance applicants to interviews, automatically rejecting the rest without human review. In its first year, the tool cut time-to-hire and significantly lightened recruiter workloads, and management called it a success. Then a rejected applicant, a woman with an advanced degree and relevant experience, filed a complaint alleging the process discriminated against women. The complaint reached the Chief Operating Officer, who was concerned about legal exposure and reputational risk and asked the analytics team to investigate. The student takes that role. The charge is deliberately open-ended so that students can do the following: determine whether the data support the complaint, explain how an objective tool could still produce a biased result, and finally recommend a set of actions based on both analysis and a defensible ethical rationale.

The dataset is an important part of the design. The recruiting system exported the 200 most recent applicants for comparable positions, one applicant per row, and Table 1 lists the fields. Two of those fields matter more than the others. The SkillAssessment is a score from a standardized test graded by trained human evaluators and is entirely independent of the AI. The AI_Score comes from TalentSift. Building the dataset with two independent measures of the same underlying quality is what makes the audit possible at all. With only the AI score, a gap in interview rates has an innocent explanation students cannot rule out, since one group might simply have been less qualified. The human-graded test removes

that explanation. When both groups score the same on it and only one group's AI scores fall, the question stops being whether women were less qualified and becomes why the tool disagreed with the humans.

Field	Description
ApplicantID	Unique identifier for each applicant
Gender	Self-reported gender (Female / Male)
Age	Applicant age in years
Department	Department applied to
YearsExperience	Years of relevant work experience
EducationLevel	Highest degree (Associate / Bachelor / Master / PhD)
SkillAssessment	Score (0 to 100) on a standardized, human-graded skills test
AI_Score	Score (0 to 100) assigned by the TalentSift AI tool
Decision	Outcome: Advanced (to interview) or Rejected

Table 1. Fields in the TechNova Applicant Dataset.

The size of the effect was adjusted carefully. The gender penalty in the AI score is large enough to push the female-to-male advance ratio below the four-fifths threshold, so students reach a finding they can defend. It is also small enough that nothing shows up from scrolling the spreadsheet alone. The disparity has to be found through analysis rather than through inspection. It is worth noting that real-world hiring algorithms rarely impose such an overt penalty, because their bias is usually spread across correlated features. This difference makes it hard to detect and prove selection bias. As this is a teaching case, we injected the penalty directly and visibly to let students learn the mechanism.

TechNova Solutions and TalentSift are fictional, the dataset is synthetic, and no field identifies or corresponds to a real person, company, or product. The data was constructed for instructional use.

4. IMPLEMENTATION

The case begins by giving students the Appendix A handout along with the dataset file. The handout carries the TechNova scenario, the COO's charge, and the six parts of the audit. Appendix B lists the Excel techniques each part requires, and Appendix C holds the instructor solution with the expected result at every step.

The case can be run in a variety of ways, depending on the course objectives and time constraints. Groups of two to four students may work through the analysis together, or each student may complete it individually and then compare results within a group. Comparing results is usually worth the extra time, because students reach different conclusions from the same spreadsheet and the differences are instructive. A typical in-class deployment takes about 45 to 60 minutes for the Excel work, preceded by a short instructor-led discussion of correlation and the four-fifths rule. The ethical analysis and the recommendation may then be assigned as follow-up work or debated in a later session. Table 2 shows a representative timeline.

Activity	Time
Correlation and the four-fifths rule (mini-lecture)	10 to 15 minutes
Introduce the four ethical perspectives	10 to 15 minutes
Excel audit (Parts 1 to 6)	45 to 60 minutes
Ethical analysis and recommendation	assigned as follow-up
Class discussion of recommendations	20 to 30 minutes

Table 2. Representative Implementation Timeline.

The order of the six parts is deliberate, and it is worth keeping. Each part either rules something out or establishes something the next part needs. Parts 1 and 2 establish that a disparity exists before any explanation is on the table, which keeps students from reasoning backward from a conclusion they already suspect. Part 3 tests the innocent explanation before the guilty one. Only then does Part 4 look for a mechanism, and only once the mechanism is located does Part 5 connect it to individuals. Students

who jump to Part 4 can find the correlation but cannot defend it, and that difference is worth surfacing when it arises.

Part 1, baseline. Using pivot tables, students describe the pool: counts of female and male applicants, counts of advanced versus rejected applicants, and the overall advance rate.

Part 2, outcomes by gender. Students compute the advance rate for women and men separately and take the ratio, then apply the four-fifths rule to check whether the ratio falls below 0.80.

Part 3, the innocent explanation. A gap in outcomes is not automatically unfair. Students test whether qualifications explain it by comparing women and men on average years of experience and on the human-graded skills test. Equal qualifications rule out the "they were less qualified" explanation.

Part 4, the mechanism. With qualifications equal but outcomes unequal, students compare the average AI_Score by gender and use CORREL to measure how the AI_Score relates to the human skill test, experience, and a numeric gender indicator. They confirm the AI score strongly predicts the decision, showing it is the channel through which the disparity operates. Instructors should expect the magnitude to mislead. The gender correlation is modest compared with the correlation between the human skills test and the AI score, and students who read size alone may conclude that the gender effect is trivial. The contrast is what matters: the human test relates to gender at essentially zero, and the AI score does not, and a modest shift is enough to move a large share of applicants across a hard cutoff.

Part 5, the individuals. Statistics describe groups; the complaint came from a person. Students filter to high-skill applicants (skills test of 70 or above) and compare advance rates for women and men within that group, then optionally use VLOOKUP to describe one qualified woman who was rejected. This turns a percentage into a person and sets up the justice analysis.

Part 6, communicate. Students build at least two charts, apply conditional formatting to the correlation results, and write a one-paragraph plain-language summary stating whether the data support the complaint and naming the mechanism.

The audit answers what happened. The COO also asked what to do, and that question has no answer in the spreadsheet. Students apply the four frameworks from Section 2.2 to what the audit found.

Utilitarian. Students weigh consequences in aggregate. The efficiency gains are real, since TalentSift reduced time-to-hire and lightened recruiter workloads. Against those gains are the harm to the applicants the tool screened out and TechNova's legal and reputational exposure. The calculation depends on whose utility is counted. If only TechNova's operating costs are counted, the tool looks successful. If the rejected applicants are counted, the result changes. Most students conclude that the harms outweigh the gains once liability and fairness are included, although a case for repairing the tool rather than retiring it may also be defended.

Deontology. A duty-based analysis asks what applicants are owed, regardless of the tool's efficiency. The relevant fact here is structural rather than statistical. Any applicant scoring below 70 is rejected automatically, with no human review. Students apply the Formula of Universal Law and ask whether TechNova could will that every employer screen applicants this way. The complainant received no individual consideration at all. This is the objection the duty-based perspective raises, and the utilitarian perspective does not.

Justice. This perspective connects most directly to the audit results. Part 5 shows that among applicants who cleared the same human skills threshold, 57.7% of women advanced compared with 91.8% of men. These are similar people receiving dissimilar outcomes, measured on a test the AI never saw. The veil of ignorance gives students a test they can apply: would you accept this screening rule without knowing which group you belong to? The four-fifths rule states the same concern in law, so a philosophical principle and a regulatory threshold reach the same conclusion.

Virtue. Virtue ethics moves the question from the act to the actor. Students ask what an honest, fair-minded manager would do after seeing a female-to-male advance ratio of 0.56 alongside identical skills

scores. Continuing to run the tool unchanged is itself a decision, and it reveals what the organization is prepared to accept. Students who treat virtue as a checklist tend to miss this point.

The case is designed in a way that the four elements of the framework do not agree automatically. Students then write a brief recommendation of roughly half a page that draws on both the audit and the ethical analysis and states the trade-off between efficiency and fairness rather than hiding it. Strong recommendations usually suspend fully automated rejection, add human review, and audit or retrain the tool against a fair benchmark such as the human skills test. Facilitation guidance appears in Appendix C.7.

Several discussion questions are listed to extend the analysis. If the tool never receives an applicant's gender, how can it discriminate on the basis of gender? Would deleting the Gender column fix the problem? Who is responsible when a purchased tool discriminates: the vendor, the HR leader who bought it, or the analyst who noticed? What type of monitoring should a company require before deploying an automated hiring tool? The first two questions matter most, because they lead students to proxy variables. A model can reconstruct a protected attribute from correlated features even when the attribute is withheld.

Assessment can follow the weight of the work. Appendix C.9 provides a 100-point rubric that assigns 75 points to the six Excel parts and 25 points to the ethical analysis and the recommendation. Instructors who run the case as a single in-class activity may instead grade participation and collect only the written recommendation. For group work, having each student submit the recommendation individually, while the team shares a single workbook, separates analytical skill from ethical judgment.

5. DISCUSSION

Students who have completed the case responded positively to it. They found the analysis simple yet informative. They liked the ethical discussion topic, which made them more vigilant and critical when using AI. That last reaction was not the stated objective, and it is interesting how the simple practice brought them a broader wariness about trusting AI output.

The core of this case is auditing an automated system's output against a benchmark, and it can serve various purposes. A database course can use it to audit a query that looks neutral but returns different outcomes across groups. In statistics, the dataset illustrates how a hard cutoff affects a small shift in a distribution: a penalty of a few points on a 0-to-100 scale pushes a substantial share of one group below the 70-point threshold. Management and ethics courses can build on the calculated results and go straight to the four-perspective decision part. The scenario is easily applicable to other high-stakes automated decisions, such as loan approval or admissions screening, by changing the outcome variable and the characteristic. The important part is that every version needs at least a two-measure design, because a proper audit requires an independent benchmark so that students can explain the gap rather than merely describe it.

However, instructors should note several limitations of the case prior to implementing it.

Synthetic data: Regenerate the data with new random values while keeping the same structure and bias mechanism; the generation logic is documented in Appendix D. The fair benchmark handed to students is a simulation of an audit rather than a real one. In practice, the model is usually a black box and no clear benchmark exists. In addition, any disparity must be detected and argued from messier

evidence. As a result, instructors must make this gap explicit, so students do not interpret real audits as being this clean.

The four-fifths rule is a heuristic: It is a screening standard the EEOC uses to flag potential adverse impact, not a legal finding of discrimination. Students should be cautioned not to overclaim from it, and stronger courses can discuss its known limitations, such as sensitivity to sample size.

Prerequisite knowledge: Students need working comfort with basic Excel and benefit from a short review of correlation. Those without it may need scaffolding on interpreting a correlation coefficient before Part 4.

Ethical vocabulary: The four-perspective analysis assumes the instructor is comfortable facilitating classical ethical frameworks. Courses that use a different ethical vocabulary can substitute their own while keeping the structure of a genuine decision under conflicting principles.

Despite these limitations, the case adapts to different class sizes, time blocks, and levels of student preparation, and its core value, giving students practice at auditing an automated decision and then judging it, holds regardless of which tool or protected characteristic the scenario uses.

6. CONCLUSIONS

This teaching case familiarizes students with algorithmic accountability through hands-on practice, without requiring any prior background in data analysis or ethics. Students detect a disparity produced by an AI job-applicant screening tool using basic Excel and trace it to the interaction between the AI screening score and the 70-point cutoff. The main takeaway is that an AI tool can be blind to gender and still discriminate, and that a data analyst can reveal this even with ordinary tools. The case also requires students to decide what the organization should do and to defend that decision under a named ethical framework. Detecting the disparity and judging it are separate skills that are essential for students entering roles where AI shapes consequential decisions.

7. REFERENCES

Ariyachandra, T. (2026). Introducing data analytics and AI collaboration to novice students: An assignment using real business data. Information Systems Education Journal, 24(6), 44-62. https://doi.org/10.62273/IJIV9938

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732. https://doi.org/10.15779/Z38BG31

Bright, D. S., Winn, B. A., & Kanov, J. (2014). Reconsidering virtue: Differences of perspective in virtue ethics and the positive social sciences. Journal of Business Ethics, 119(4), 445-460. https://doi.org/10.1007/s10551-013-1832-x

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications, 10, Article 567. https://doi.org/10.1057/s41599-023-02079-x

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

Firth, D. R., & Triche, J. (2024). Generative AI in practice: A teaching case in the introduction to management information systems class. Information Systems Education Journal, 22(4), 29-47. https://doi.org/10.62273/LDVL8354

Gustafson, A. (2013). In defense of a utilitarian business ethic. Business and Society Review, 118(3), 325-360. https://doi.org/10.1111/basr.12013

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy, 63(4), 309-321. https://doi.org/10.1086/257678

Ip, E. (2025). Fair AI in hiring: Experimental evidence on how biased hiring algorithms and different debiasing methods affect the quality and diversity of applicants. Behavioral Science & Policy, 11(1), 44-54. https://doi.org/10.1177/23794607251353585

Kim, P. T. (2017). Data-driven discrimination at work. William & Mary Law Review, 58(3), 857-936. https://scholarship.law.wm.edu/wmlr/vol58/iss3/4

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context. Business & Information Systems Engineering, 63(1), 39-54. https://doi.org/10.1007/s12599-020-00673-w

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research, 13(3), 795-848. https://doi.org/10.1007/s40685-020-00134-w

Larson, B. E., Bohler, J. A., & Bolekar, N. (2026). Developing critical thinking in data analytics education: A teaching case evaluating ChatGPT responses to a visualization. Information Systems Education Journal, 24(6), 31-43. https://doi.org/10.62273/RLDU1198

Njoto, S., Cheong, M., Frermann, L., & Ruppanner, L. (2025). Bias and discrimination against women and parents in semi-automated hiring systems. New Technology, Work and Employment. https://doi.org/10.1111/ntwe.12321

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In Proceedings of the 2020 Conference on Fairness,

Accountability, and Transparency (pp. 469-481). Association for Computing Machinery. https://doi.org/10.1145/3351095.3372828

Rawls, J. (1971). A theory of justice. Harvard University Press.

Scharding, T. K. (2015). Imprudence and immorality: A Kantian approach to the ethics of financial risk. *Business Ethics Quarterly, 25*(2), 243-265. https://doi.org/10.1017/beq.2015.17

Sharp, J. H., Anderson, J. E., & Lang, G. (2026). AI in IS education: A categorization of ISCAP publications using Zuboff's automate-informate-transformate framework. Information Systems Education Journal, 24(4), 22-35. https://doi.org/10.62273/QGPE2646

Sison, A. J. G., & Ferrero, I. (2015). How different is neo-Aristotelian virtue from positive organizational virtuousness? Business Ethics: A European Review, 24(S2), S78-S98. https://doi.org/10.1111/beer.12099

U.S. Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. 29 C.F.R. § 1607. https://www.eeoc.gov/laws/guidance/uniform-guidelines-employee-selection-procedures

World Economic Forum. (2023, May 3). Growth Summit 2023: Job creation and reskilling must be central to growth in the age of uncertainty, advancing AI and the green transition [Press release]. https://www.weforum.org/press/2023/05/growth-summit-2023-job-creation-and-reskilling-must-be-central-to-growth-in-the-age-of-uncertainty-advancing-ai-and-the-green-transition/

APPENDIX A

Student Assignment (Handout)

This appendix is the student-facing assignment. Instructors may distribute it directly or adapt it. It corresponds to the analysis described in the main manuscript.

Note on the data: TechNova Solutions and its TalentSift screening tool are fictional, created for this exercise. The applicant dataset is synthetic and does not describe any real people or company. It was built so you can practice auditing an automated decision using realistic-looking data.

Background

You are a business analyst at TechNova Solutions, a technology-services firm of about 900 employees. Facing a flood of job applications, TechNova bought an AI resume-screening tool called TalentSift. TalentSift gives each applicant a score from 0 to 100. Anyone scoring 70 or above is advanced to an interview; everyone else is rejected automatically, with no human review. In its first year, the tool reduced hiring time and lightened recruiters' workloads, and management was pleased.

Recently, a rejected applicant, a woman with an advanced degree and several years of experience, filed a complaint. She was rejected without an interview and questioned whether an automated system can be fair. The complaint reached the Chief Operating Officer (COO), who has asked your team to investigate.

Your charge from the COO: Determine whether the data support the complaint. If they do, explain how a tool built to be objective could produce a biased result. Then recommend what TechNova should do, grounded in both the numbers and a defensible ethical rationale.

The Data

You are given TechNova_Hiring_Data.xlsx, containing the 200 most recent applicants for comparable positions (one row per applicant). The Data Dictionary tab describes each field. Two fields matter most: SkillAssessment, a human-graded skills test unrelated to AI, and AI_Score, the score assigned by TalentSift. Comparing these two is the key to the investigation.

Part 1: Establish a Baseline (Pivot Tables)

Before comparing groups, describe the applicant pool. Using pivot tables, find:
1. The number of female and male applicants.
2. The number of applicants Advanced versus Rejected overall.
3. The overall advance rate (percentage Advanced).

Hint for computing a rate: create a helper column that equals 1 for "Advanced" and 0 for "Rejected," then use the average of that column.

Part 2: Compare Outcomes by Gender (Four-Fifths Rule)

Build a pivot table showing the advance rate for women and men separately. Then:
1. Report the percentage of women advanced and the percentage of men advanced.
2. Compute the ratio of the female advance rate to the male advance rate.
3. The four-fifths rule says that if one group's selection rate is less than 80% (0.80) of the highest group's, the difference may indicate adverse impact. Does your ratio fall below 0.80?

Part 3: Check Whether Qualifications Explain the Gap

A gap in outcomes is not automatically unfair; maybe one group was more qualified. Test that directly. Using pivot tables, compare women and men on the two qualification measures that do NOT come from the AI:
1. Average YearsExperience by gender.

2. Average SkillAssessment (the human-graded test) by gender.

If the two groups are similar on both, then qualifications cannot explain the outcome gap. What do you find?

Part 4: Locate the Mechanism (CORREL)

If qualifications are equal but outcomes are not, something in between is driving the result. Investigate the AI score:
1. Compare the average AI_Score for women and men using a pivot table.
2. Create a numeric gender column (1 for Female, 0 for Male). Use CORREL to measure the relationship between AI_Score and each of SkillAssessment, YearsExperience, and the gender indicator. Which is the AI score most related to? Is it related to gender even though the human skills test is not?
3. Create a numeric decision column (1 for Advanced, 0 for Rejected) and use CORREL to confirm the AI_Score strongly predicts the Decision.

Part 5: Find the Individuals Affected (Filtering, VLOOKUP)

Statistics describe groups, but the complaint came from a person. Filter to high-skill applicants (SkillAssessment of 70 or above) and compare the advance rate for women and men within that group. Among applicants who clearly had the skills, how differently were women and men treated? Optionally, use VLOOKUP to look up one specific ApplicantID and describe a qualified woman who was rejected.

Part 6: Visualize and Summarize

Create at least two charts (for example, advance rate by gender and average AI_Score by gender). Apply conditional formatting to your correlation results to highlight the strongest relationships. Then write a short, plain-language summary (one paragraph) stating whether the data support the complaint and naming the mechanism responsible.

Part 7: The Ethical Decision

Your numbers are only half of what the COO wants. Analyze the situation through four ethical perspectives, then recommend an action.
Utilitarian: Which action produces the greatest overall benefit and least harm across applicants, employees, and the company? Does the tool's efficiency outweigh the harm to rejected applicants?
Deontology: Do applicants have a right to fair, non-discriminatory consideration? Does automated rejection with no human review respect or violate that right?
Justice: Are similar people treated similarly? Your Part 5 finding is directly relevant.
Virtue: What would an honest, fair-minded manager do upon seeing these numbers? What does continuing to use the tool unchanged say about the organization's character?

Part 8: Proxy Variables

In Part 4 you found the AI score is related to gender even though the human skills test is not. Explain how the tool can discriminate by gender when it never receives an applicant's gender, and why removing the Gender column alone would not make it fair.

Part 9: Recommendation to the COO

Write a brief recommendation to the COO, about half a page. State clearly what TechNova should do about TalentSift and why. A strong recommendation names the trade-off between efficiency and fairness instead of hiding it, and it weighs options such as suspending the tool, adding human review, auditing or retraining it, or changing the score threshold.

Additional Discussion Questions

1. Who is responsible when a purchased AI tool discriminates: the vendor, the HR leader who bought it, or the analyst who noticed?

ISSN: 2473-4901
v12 n6957

2. What ongoing monitoring should a company put in place before deploying any automated hiring tool?

APPENDIX B

Excel Techniques Used in This Case

This appendix lists the introductory Excel techniques the case requires. Instructors should insert annotated screenshots from their own Excel version, since menu layouts differ across releases and platforms. Each technique below notes where it is used in the assignment.

Technique	Where used	What students do
PivotTable	Parts 1, 2, 3, 4	Count applicants and compute advance rates and average scores by gender.
Helper column (0/1 coding)	Parts 1, 2, 4	Convert Advanced/Rejected and Female/Male into numbers so they can be averaged and correlated.
CORREL function	Part 4	Measure the relationship between AI_Score and skill, experience, gender, and the decision.
Filtering	Part 5	Isolate high-skill applicants (SkillAssessment >= 70).
VLOOKUP	Part 5	Look up one applicant's full record by ApplicantID.
Charts (bar)	Part 6	Show advance rate by gender and average AI_Score by gender.
Conditional formatting	Part 6	Highlight the strongest correlations in the results table.

Table B1. Excel Techniques Mapped to Assignment Parts.

Step-by-step instructions for the main techniques (Excel for Mac, Microsoft 365, 2025)

PivotTable (advanced rate by gender). First, add a helper column named AdvNum with the formula =IF(Decision="Advanced",1,0), so each row is 1 for Advanced and 0 for Rejected. Select the data range including the headers. On the ribbon, click Insert > PivotTable, keep the selected range, choose New Sheet, and click OK. In the PivotTable Fields pane, drag Gender to Rows and drag AdvNum to Values. The value shows as "Sum of AdvNum"; click that field in the Values area, open Field Settings (the small i info button next to the field on Mac, or Control-click the field and choose Field Settings), and set Summarize by to Average. The result is the advance rate for each group.

CORREL. First, make a numeric gender column with =IF(Gender="Female",1,0). In an empty cell type =CORREL (array1, array2), for example =CORREL (F2:F201, G2:G201), where F is AI_Score and G is the numeric gender column. The result runs from -1 to 1; a value near 0 means no relationship.

VLOOKUP. Type =VLOOKUP (lookup_value, table_range, column_number, FALSE). Example: =VLOOKUP ("A1011", A2:H201, 5, FALSE) returns the value in the 5th column for applicant A1011. Use FALSE for an exact match.

Suggested screenshots to include: (a) the PivotTable Fields pane set up for advance rate by gender; (b) a CORREL formula in the formula bar with its result; (c) the filter dialog set to SkillAssessment >= 70; (d) a completed bar chart of advance rate by gender; and (e) a correlation table with conditional formatting applied.

APPENDIX C

Instructor Solution, Expected Results, and Grading Rubric

These are instructor materials. Do not distribute to students. All figures below are drawn from the provided dataset (TechNova_Hiring_Data.xlsx; 200 applicants: 98 female, 102 male). Minor rounding differences are acceptable, and figures will change if the dataset is regenerated (see Appendix D).

C.1 Baseline (Part 1)

Applicants: 98 female, 102 male (200 total).
Overall: 91 Advanced, 109 Rejected, an overall advance rate of about 45.5%.

C.2 Outcomes by Gender (Part 2)

Gender	Advance Rate	Interpretation
Female	32.7%	Well below male rate
Male	57.8%	Highest group
Ratio (F / M)	0.56	Below 0.80, adverse-impact flag

Table C1. Advance Rate by Gender.

The female-to-male ratio of about 0.56 is far below the four-fifths (0.80) threshold, so the rule is triggered. This is the first sign the complaint may have merit.

C.3 Is the Gap Justified by Qualifications? (Part 3)

Measure	Female	Male
Avg. YearsExperience	5.2	5.2
Avg. SkillAssessment (human test)	70.5	70.5

Table C2. Qualifications by Gender.

On both non-AI measures, women and men are essentially equal. A difference in qualifications cannot explain the gap in advance rates, so the "they were just less qualified" explanation is not supported.

C.4 The Mechanism (Part 4)

Average AI_Score: Female 64.7, Male 71.2, a gap of about 6.5 points despite equal human skill scores.

Variable	Correlation with AI_Score	Reading
SkillAssessment	0.79	Strong, as expected
YearsExperience	0.62	Moderate to strong, as expected
Gender (Female = 1)	-0.27	Being female lowers the AI score

Table C3. Correlations with the AI Score.

The key teaching point: the human SkillAssessment has essentially no relationship to gender (correlation about 0.00), yet the AI_Score is negatively related to being female (about -0.27). The AI has introduced a gender relationship that the fair, human-graded measure does not contain. Students should also confirm that the AI_Score strongly predicts the Decision (correlation about 0.81), so the biased score directly drives who gets an interview.
The audit uses descriptive statistics only: rates, ratios, and correlations, which suits an introductory audience. Instructors teaching a more quantitative section can add a chi-square test on the advance-by-gender counts and a two-sample t-test on AI_Score by gender as an extension.

C.5 Individuals Affected (Part 5)

Gender	Advance Rate (high-skill only, SkillAssessment >= 70)
Female	57.7%
Male	91.8%

Table C4. High-Skill Advance Rate by Gender.

Among applicants who demonstrably had strong skills (52 women and 49 men in this dataset), nearly all the men advanced, while more than four in ten qualified women were rejected. A concrete VLOOKUP example: applicant A1011 is a woman who scored 81.7 on the human skills test, well above the bar, but received an AI_Score of 64 and was rejected. The complaint came from exactly this group, which makes the justice analysis concrete.

C.6 Model Summary Students Should Reach

A strong summary reads something like: the data support the complaint. Women and men were equally qualified on experience and on a human-graded skills test, yet the AI assigned women lower scores, and because the AI score determined interviews, women advanced far less often. The gap fails the four-fifths rule, and even highly skilled women were rejected at high rates. The AI score is the mechanism producing the disparity.

C.7 Facilitating the Ethical Analysis

Students often want a single correct answer. Emphasize that the four frameworks can pull in different directions and that a strong answer weighs the tension rather than hiding it.

Utilitarian: The tool creates real efficiency (lower cost, faster hiring), but the harm falls heavily on one group and includes legal and reputational risk. Many students conclude the harms outweigh the benefits once fairness and liability are counted; a well-argued "fix the tool rather than abandon it" case is also acceptable.

Deontology: Applicants arguably have a right to fair, non-discriminatory consideration. Automated rejection with no human review is the sharpest rights concern.

Justice: The Part 5 result, equally skilled people treated unequally, is a textbook justice violation and is usually the most decisive perspective for students.

Virtue: Ask what a manager of good character does on seeing these numbers. Continuing unchanged is hard to defend as honest or fair, which frames inaction as itself a choice.

C.8 Proxy Variables (Part 8)

The tool never sees gender, yet it scores women lower because it learns proxies. Features such as word choice, employment gaps, and the schools an applicant attended correlate with gender in the training data, so the model reconstructs a gender signal from them.
Deleting the Gender column does not remove those proxies, so it does not remove the bias. This is the central conceptual point of the case, and it is why a real fix means auditing or retraining the tool against a fair benchmark such as the human skills test, not just hiding the protected attribute.

C.9 Recommendation (Part 9)

A strong recommendation pauses fully automated rejection, adds human review, and audits or retrains the tool against a fair benchmark such as the human skills test, while acknowledging the efficiency cost. Give credit to students who state the efficiency-fairness trade-off directly rather than writing around it.

C.10 Notes on the Discussion Questions

Responsibility is distributed: the vendor built the tool, the HR leader deployed it without an audit, and the analyst who noticed now has a duty to report. There is no single correct answer, which is the point. Reasonable monitoring includes pre-deployment bias testing, ongoing four-fifths checks, human review of borderline cases, and vendor audit requirements.

C.11 Suggested 100-Point Grading Rubric

Component	Points	What to look for
Baseline & outcomes (Parts 1-2)	20	Correct pivot tables; correct advance rates and F/M ratio; applies the four-fifths rule.
Qualification check (Part 3)	15	Recognizes equal qualifications; correct interpretation.
Mechanism / correlations (Part 4)	20	Correct CORREL results; identifies the AI score as the biased driver.
Individual / high-skill analysis (Part 5)	10	Correct filtered comparison; connects the result to the complaint.
Charts, formatting, summary (Part 6)	10	Clear charts; accurate plain-language summary.
Ethical analysis (Part 7)	10	All four frameworks applied substantively.
Proxy-variable reasoning (Part 7)	5	Identifies proxy variables; explains why removing Gender does not fix the bias.
Recommendation (Part 7)	10	Defensible; acknowledges the efficiency-fairness trade-off.

Table C5. Suggested Grading Rubric (100 Points).

APPENDIX D

Dataset Generation Logic (Reproducibility and Answer-Sharing Control)

Instructors note: Do not share Appendix D or the generator script with students. It reveals the injected bias. Keep it in an instructor-only location, separate from the dataset file.

The dataset is synthetic. It was built so that qualifications are balanced across gender while the AI score carries a systematic penalty against female applicants that is large enough to fail the four-fifths rule but not obvious on inspection. The logic below lets an instructor regenerate an equivalent dataset with new random values, which prevents answer-sharing across semesters while preserving every teaching point. The provided file uses a fixed random seed so that the figures in Appendix C are reproducible. The penalty is set to 6.5 points on the 0-to-100 scale, the smallest value that reliably pushes the female-to-male advance ratio below the 0.80 threshold given the 70-point cutoff. It is a teaching parameter, not an estimate of any real hiring system.

D.1 Construction Rules

1. Create 200 applicants, 98 female and 102 male, in random order.
2. Draw YearsExperience and SkillAssessment from the same distributions for both genders, then mean-match each measure so the female and male averages are equal. This guarantees the two groups are equally qualified by construction.
3. Compute the AI_Score as a function of standardized SkillAssessment (strong positive weight), standardized YearsExperience (moderate positive weight), a female indicator (a negative penalty, the injected bias), and random noise. The human SkillAssessment contains no gender term; only the AI_Score does.
4. Set Decision to Advanced when AI_Score is 70 or above; otherwise, Rejected.

Increasing the size of the female penalty widens the disparity; decreasing it makes the bias subtler. The provided file uses a penalty that yields a female-to-male advance ratio of about 0.56.

D.2 Reference Implementation (Python)

```
import numpy as np, pandas as pd
rng = np.random.default_rng(20260713)   # change seed for a fresh dataset
N, nF, nM = 200, 98, 102
g = np.array(['Female']*nF + ['Male']*nM); rng.shuffle(g)

df = pd.DataFrame({'Gender': g})
df['YearsExperience'] = np.clip(rng.normal(5.2, 2.2, N), 0, 15).round(1)
df['SkillAssessment'] = np.clip(rng.normal(70.5, 11, N), 20, 100).round(1)

# make qualifications exactly equal across gender
for col, tgt in [('SkillAssessment',70.5), ('YearsExperience',5.2)]:
    for gg in ['Female','Male']:
        mask = df.Gender==gg
        df.loc[mask,col] = df.loc[mask,col] - df.loc[mask,col].mean() + tgt

fem = (df.Gender=='Female').astype(int).values
zs = (df.SkillAssessment - df.SkillAssessment.mean())/df.SkillAssessment.std()
zy = (df.YearsExperience - df.YearsExperience.mean())/df.YearsExperience.std()

ai = 71.5 + 8.5*zs + 6.0*zy - 6.5*fem + rng.normal(0,3.2,N)
df['AI_Score'] = np.clip(ai.round(0), 0, 100).astype(int)
df['Decision'] = np.where(df.AI_Score>=70, 'Advanced', 'Rejected')
```

Add ApplicantID, Age, Department, and EducationLevel as descriptive fields (they do not affect the outcome). The full generator and the resulting workbook are provided with the supporting materials.