



Journal of Philosophical Investigations

Journal of Philosophical Investigations

Print ISSN: 2251-7960 Online ISSN: 2423-4419

Homepage: <https://philosophy.tabrizu.ac.ir>



University of Tabriz

Algorithmic Counter-Authority and the Epistemic Legitimacy of Large Language Model Advice: Gendered Power in Pakistan as a Philosophical Test Case

Hafiz Muhammad Sikandar^{✉1} | Lihui Liu²

1. Corresponding Author, PhD Candidate in English Language and Literature, College of International Studies, Southwest University, China. E-mail: sikandarghouri999@gmail.com.
2. Professor of English and American Literature, College of International Studies, Southwest University, China. E-mail: liulihui99@163.com.

Article Info

Article type:

Research Article

Article history:

Received 03 August 2026
Received in revised form
10 August 2026
Accepted 11 August 2026
Published online 25
October 2026

Keywords:

Islamic Feminism
Gendered Authority
Epistemic Injustice AI
epistemology

ABSTRACT

Large language models no longer only supply information; they increasingly render advice, weighing considerations and favoring one course of action, thereby making an implicit claim to some standing to counsel at all. This article treats that shift as a philosophical problem. It develops algorithmic counter-authority to describe how AI-generated responses can limit, reinterpret, or displace an established claim to gendered authority through the reasons they give and the sources they invoke. The concept is developed through close analysis of four AI-generated responses to dilemmas drawn from Pakistani family life: a father's claim over a daughter's university education, a husband's claim to end his wife's postgraduate study, and a father's and brothers' claim to restrict a woman's public employment. These cases serve as philosophical test material, not an empirical sample of provider behavior, and no claim below concerns how any system typically behaves. The central argument is that a system can perform counter-authority convincingly while lacking the epistemic and moral standing ordinarily expected of an adviser: situated knowledge, relational accountability, first-person understanding, recognized religious credential, and accountability for what follows from its counsel. This produces a paradox distinctive to artificial advice-giving: an output can be emancipatory in content while remaining structurally unaccountable in its delivery. Contributing to debates on AI epistemology, artificial moral agency, alignment, and technological power, the article shows that rhetorical confidence did not reliably track evidentiary adequacy across the cases and asks whether this deficit is unique to artificial systems or only intensified and concealed by them.

Cite this article: Sikandar, M. H. & Liu, L. (2026). Algorithmic Counter-Authority and the Epistemic Legitimacy of Large Language Model Advice: Gendered Power in Pakistan as a Philosophical Test Case. *Journal of Philosophical Investigations*, 20(56), 337-356. <https://doi.org/10.22034/jpiut.2026.73945.4655>



© The Author(s).

Publisher: University of Tabriz.

Introduction

A search engine, asked whether a nineteen-year-old woman should accept a university place her father opposes, returns a ranked list of pages and leaves the weighing to her. A conversational AI system, put the same question, returns something else: a composed answer, in the register of counsel, that selects among considerations, states a position, and tells her what she should do. Advice is philosophically distinct from information in ways that matter before any question of accuracy arises. It selects which considerations are relevant and which may be set aside; it ranks possible actions rather than listing them; it presents reasons in a sequence built to support a conclusion; and in doing this it positions the system offering it as entitled, in some sense, to guide conduct. An adviser, whatever else it is, makes a claim to standing simply by advising. This article asks what that claim amounts to when the adviser is a large language model and the conduct in question concerns a woman's education, marriage, or public employment under conditions of contested family authority.

Locating these dilemmas requires some sense of the terrain on which they are staged. Pakistan's 1973 Constitution already supplies the equality a reader unfamiliar with the country might expect AI counsel to import from outside it: Article 25 prohibits discrimination on the basis of sex, Article 25-A makes education a right unqualified by sex, and Article 34 commits the state to securing women's full participation in national life (Constitution of the Islamic Republic of Pakistan, 1973, arts. 25, 25-A, 34). The distance between that text and its exercise registers internationally: the World Economic Forum's 2025 Global Gender Gap Report ranks Pakistan last among the 148 countries it surveys, and women's share of the labour force sits near a fifth, among the lowest recorded anywhere; the one dimension in which the country's score improved was educational attainment, and even that owes partly to falling male enrolment rather than to any rise in women's own participation ([World Economic Forum, 2025](#)). A national household survey found that two in five women needed a family member's permission to take up or remain in paid employment, most citing a husband's or father's refusal rather than their own preference ([Minardi et al., 2021](#)); contemporary reporting confirms the pattern from inside a single household, describing women who still require family permission to work outside the home even as economic pressure pushes more of them to try anyway ([Agence France-Presse, 2024](#)). Marriage carries a comparable gap between statute and settlement: a 2025 law raised the minimum marriage age in the capital territory to eighteen for both sexes and criminalized facilitating an underage marriage, only for the Council of Islamic Ideology, the constitutional body reviewing legislation for conformity with Islam, to declare the measure un-Islamic on the ground that puberty rather than a fixed age should govern marriageability, a position now argued before the Federal Shariat Court ([Shabbir, 2025](#)). Fathers, husbands, and brothers do not draw the authority a woman in these dilemmas confronts from one settled source; they draw it from a contest among sources still visibly unresolved at the level of the state itself.

The question is not new in kind. Human advisers, too, sometimes claim more standing than they possess, and philosophers have long distinguished the epistemic authority a person accrues through demonstrated reliability from authority merely performed through confident manner. What is new is the scale and texture of the performance now available to a system that has never occupied a role in anyone's household and bears no responsibility for what happens after it finishes speaking. This article examines that performance using AI-generated responses to three closely related dilemmas of gendered authority in Pakistan as test material: a father who tells his admitted daughter that a respectable woman remains at home; a husband who asks his wife to abandon a postgraduate degree that does not, on her account, interfere with her domestic responsibilities; and a father and brothers who treat a woman's online teaching work as a matter of family honor rather than personal choice. These are cases in which several sources of authority compete for the same decision, familial precedent, religious interpretation, constitutional right, and the woman's own judgment, and a system asked to adjudicate among them cannot avoid taking a position on which source should prevail, even when it declines to say so explicitly.

Existing scholarship on AI and gendered dilemmas of this kind tends to examine bias, factual accuracy, user trust, or persuasive effect, but rarely asks what has to be true of a system for its advice to be legitimately action-guiding, independent of whether the advice happens to be accurate or the recipient happens to trust it. A response can be factually correct and normatively appealing while lacking the standing an adviser on a matter this intimate would ordinarily be expected to have, since whether the content is right and whether the voice is entitled to deliver it are logically separable questions, even when a fluent register makes them feel like one. This separation is the philosophical gap the article works to keep open.

The concept developed to hold that gap open is algorithmic counter-authority: the capacity of AI-generated advice to question, delimit, reinterpret, or redirect an established claim to gendered authority by presenting alternative reasons, interpreters, or courses of action in its place. The term is deliberately double-edged: counter-authority names something a response does, discursively, not a standing the system thereby acquires. A system can perform it with real skill while remaining unentitled, in every respect that would normally ground an adviser's legitimacy, to the position it has assumed. The thesis defended below is that large language models can perform algorithmic counter-authority while lacking the situated knowledge, relational accountability, first-person understanding, recognized religious credentials, and accountability for consequences that would convert performance into possessed authority. The resulting paradox: an LLM may challenge one form of epistemic injustice, the discounting of a woman's claim to equal standing, while reproducing another, the discounting of her situated knowledge of her own circumstances, through confident, decontextualized, and institutionally unaccountable counsel.

Three questions organize the analysis that follows: what form of authority do large language models perform when they challenge paternal, marital, religious, or culturally inherited claims to obedience, and through what rhetorical means; can a system lacking situated understanding, relational accountability, and moral responsibility possess sufficient epistemic or moral standing to guide action rather than merely inform it; and how should apparently emancipatory AI advice be evaluated when its normative content may resist gendered domination even as its delivery conceals the system's own legitimacy deficit? These are philosophical questions, examined through close reconstruction of a small number of textual cases, not claims about provider behavior (the scope and non-representativeness of the case material is set out in §4). The Pakistani dilemmas remain central not as the article's ultimate subject but as a context specific enough to make the underlying structure visible. The wider stakes concern the epistemology of artificial intelligence, the moral standing of artificial agency, the pluralism value alignment is meant to respect, and the technological power a system exercises by being consulted on questions it has no standing, in the traditional sense, to answer.

1. From Algorithmic Output to Action-Guiding Authority

Why should a chatbot's response to a family dispute be studied as a performance of authority at all, rather than simply as text that happens to be persuasive or unpersuasive? The answer requires distinguishing several things that a fluent output can easily blur together: causal influence, the capacity of a response to influence what someone does next, whether or not this study observes any particular reader acting on it; rhetorical authority, the impression of competence a response creates through its manner of address; and epistemic authority, a warranted claim to know something relevant better than the person being advised. Two further terms complete the set: moral authority, a warranted claim to judge what someone ought to do, and socially recognized legitimacy, the belief that the one issuing a directive is entitled to do so. A system can possess the first two without the last three, and the claim this article develops is that current large language models occupy exactly that position when they advise on gendered family dilemmas: potentially causally influential, rhetorically confident, and lacking the situated standing that would convert either into epistemic or moral authority proper.

It follows that a sentence like the model recommends she assert her right to continue studying should be read carefully. The system does not, in any philosophically defensible sense, recommend in the way a trusted aunt or an imam recommends, drawing on a relationship, a track record, or an institutional office; it performs a recommendation, producing the surface features, evaluative vocabulary, ranked considerations, an explicit practical conclusion, that a genuine recommendation would have, without the standing that would make the performance more than a performance. It projects authority through the register of settled counsel; enacts it through numbered structure, confident verbs, and citation; and simulates deliberative competence by working through considerations in the sequence a careful adviser would use, without undergoing anything like deliberation. None of these licenses the stronger claim that

the model knows the woman's situation, believes one course of action is right, or decides on her behalf; those verbs import a first-person interiority the evidence does not warrant, and using them unqualified would concede by vocabulary what the argument means to keep open by analysis.

The performance examined here operates at the textual and rhetorical level, not at the level of demonstrated social uptake. Nothing in this article establishes that any woman who reads such a response comes to believe the system has the standing its prose assumes: a performance of authority can be shown to outrun the standing available to sustain it independent of whether any particular reader is taken in by it, which marks off a question philosophy is equipped to answer, whether a performance is warranted by the standing behind it, from a question only fieldwork could answer, whether anyone was actually persuaded.

This connects to a wider, familiar dependence on algorithmic judgment, what [DanaHER \(2016\)](#) calls the threat of algocracy: governance by algorithmic systems that reduces human participation in, and comprehension of, decisions that increasingly rely on their output. A user who asks a search engine for information retains the work of weighing what she finds; a user who asks a conversational system for advice on a family conflict has already delegated part of that weighing to a system with no personal relationship to her, no consequence riding on what follows, and no presence in the conversation her answer is meant to help her have. The delegation is not obviously a mistake; a woman with no other counsel available may reasonably prefer a confidently reasoned answer to none. What deserves scrutiny is not the choice to consult such a system but its own silence about what it is: none of the performative markers described discloses that a delegation has occurred, or that the receiving party cannot be held to it afterward. That silence, rather than the advice's content, is where the philosophical problem of algorithmic counsel resides, and it is the problem the next section is built to make precise.

2. Philosophical Framework

2-1 Legitimacy and the performance of authority

Weber's account of authority begins from a distinction worth keeping sharp: authority is not the same as power, because authority additionally requires belief in legitimacy on the part of those who submit to it ([Weber, 2019, 339](#)). A father's claim that a respectable daughter must obey him draws on what Weber calls traditional authority, legitimacy inherited from role and convention. A large language model has no access to this kind of legitimacy, inheriting no role or position in anyone's family; what it can perform instead is closer to Weber's legal-rational mode, the impersonal, rule-governed style he associates with bureaucratic administration ([Weber, 2019, 344, 350](#)), without thereby being neutral. A rationalized style is itself a rhetorical resource, and [Campbell's \(2007\)](#) four-layer account of religious authority, hierarchy, structure, ideology, text, sharpens the point: a system with no place in the hierarchy can still perform authority through the other three channels, reproducing an ideological position or citing scripture in a manner that reads as settled rather than contested.

[Cheong \(2017, 2021\)](#) develops a parallel point from communication practice, describing religious authority as continuously co-produced through mediated interaction rather than possessed once and exercised thereafter, and naming the relevant case bounded religious automation, systems that communicate a version of authority at work in spiritual contexts despite limits users may not see clearly. [Lustig and Nardi's \(2015\)](#) concept of algorithmic authority, legitimate power residing not in code alone but in a diversity of sociotechnical actors who validate and route around it, transfers from its original domain, trust in cryptocurrency systems, to the considerably higher-stakes register examined here.

Whether users actually extend trust of this kind to AI-generated moral advice is a question this article does not answer empirically; [Logg, Minson, and Moore \(2019\)](#) found laypeople sometimes weight advice more heavily once they believe it came from an algorithm, while [Dietvorst, Simmons, and Massey \(2015\)](#) show this trust collapses once an algorithm is observed to err. Whether this extends to character and ethical judgment, rather than the numeric forecasting tasks the original studies examined, remains open and is bracketed here. What the wider empirical literature does establish is that a model's moral output is not neutral with respect to whose values it encodes: [Abdulhai et al. \(2024\)](#) show models express identifiable, prompt-sensitive moral-foundations profiles rather than a single stable one; [Santurkar et al. \(2023\)](#) found model-elicited opinions did not consistently track the demographic groups they were meant to represent; and [Zewail et al. \(2026\)](#) find models systematically underweight moral foundations salient to populations furthest from the Western, educated, industrialized, rich, and democratic profile their training data over-represents. [Ramezani and Xu \(2023\)](#) and [Aksoy \(2025\)](#) add that English- and multilingual models alike capture cross-country moral variation poorly, bearing directly on any English-language exchange concerning Pakistani family life.

2-2 Government of conduct and algorithmic counsel

Where Weber asks what makes a command legitimate, Foucault asks how conduct gets administered without anyone necessarily issuing a command, through what he calls the production of truth ([Foucault, 1991b, 79](#)); the family, on this account, functions as an instrument of government rather than a model for it ([Foucault, 1991a, 99](#)). [Rouvroy and Berns \(2013\)](#) extend this to computational systems specifically, describing algorithmic governmentality as a mode of power that models, anticipates, and pre-emptively affects possible behaviors, and that creates a reality at least as much as it records one. A chatbot responding to a woman's question about her father's opposition does not command her to do anything; it structures what appears as a reasonable, responsible, or risky course of action, which gives the analysis a question Weber's framework cannot ask on its own, not on what authority a response speaks, but what range of conduct it makes thinkable, and what it forecloses by never mentioning it.

A response that frames her options as obedience or estrangement governs her conduct differently than one that frames them as obedience, negotiated compromise, or an appeal to law,

even when both reach the identical conclusion. The range of thinkable action, not only the verdict reached, is itself an exercise of the kind of power Foucault has in mind, whether or not the system or its designers intended it. [Kandiyoti's \(1988\)](#) concept of the patriarchal bargain, women under classic patriarchy negotiate within male control rather than simply complying with or resisting it, is put to a different use here: applied not to women's own strategies but to what AI-generated advice recommends those strategies be. [Agha \(2021\)](#) and [Khan and Biao \(2025\)](#) give the concept empirical depth in the Pakistani setting; [Purewal and Hashmi \(2015\)](#) and [Moghadam \(1992\)](#) caution that what varies across restrictive households is content rather than the presence of male authority, and that this content owes as much to kinship structure and uneven modernization as to Islam; [Zulfiqar and Kuskoff \(2024\)](#) add that education does not convert automatically into empowerment. §5.1 returns to this distinction, since several responses examined below turn out to counsel a bargain rather than a rupture while their headline conclusion sounds like rupture.

2-3 Epistemic injustice and situated understanding

[Fricker's \(2007\)](#) account of epistemic injustice adds a dimension neither Weber nor Foucault reaches alone, asking not what authorizes a claim but whether the person it is about is recognized as a credible knower of her own situation. Testimonial injustice, prejudice discounting what someone says, and hermeneutical injustice, lacking the interpretive resources to make one's own experience intelligible because one participates unequally in producing social meanings ([Fricker, 2007, 1, 6](#)), name two distinct risks a single AI response can run simultaneously. A response can supply a woman with concepts, unequal standard, coercion, conditional obedience, that expand what she can say about her situation, a hermeneutical gain in Fricker's sense; the same response can pronounce on her specific circumstances with a confident, procedural register regardless of what it actually knows about her household, finances, or the people she would negotiate with, discounting her situated understanding for its own generic account. The two effects do not cancel out; §5 argues that this double movement, rather than either effect alone, is the more precise description of what several of the reconstructed cases below actually do.

The stakes of this pairing are specific to artificial advice in a way they are not for a human adviser who errs. A human counselor who overgeneralizes can be corrected, held to account, or known to have limited insight into a family she has never met; a system performing the same overgeneralization offers no equivalent signal, since its confident register is uniform whether or not it has anything specific to go on, and the reader has no independent way of telling which is which.

2-4 Islamic feminist hermeneutics and interpretive plurality

Qur'anic verse 4:34, and specifically the term *qawwamun*¹, is the textual hinge on which most disputes over male authority in Muslim family life turn, and a system advising on any of the three dilemmas below is almost certain to encounter it. [Wadud \(1999\)](#) argues that the classical rendering of *qawwamun* as unconditional male guardianship imports seventh-century Arabian custom into a text she reads as committed to spiritual equality; [Barlas \(2019\)](#) insists on a sharp distinction between the Qur'an as revealed text and *fiqh*², the historically situated human interpretation built around it, arguing that patriarchal exegesis has been read back into the text as though the two were identical; [Mir-Hosseini \(2006\)](#) makes the equivalent move at the level of method, distinguishing Shari'a as divine from *fiqh* as fallible human derivation. None of the three asks the reader to set religion aside; each contests a monopoly on interpretation from a position inside the tradition, which is also how [Al-Sharmani \(2013\)](#) frames Islamic feminist scholarship generally. The three readings are not, however, methodologically identical: Wadud reconstructs the authorial intent behind a contested term, while Barlas and Mir-Hosseini locate the same disagreement at the level of who is authorized to interpret rather than what the term originally meant, a divergence occasionally read within the scholarship as itself a disagreement about how much weight authorial intent can bear. The distinction between revelation and its historically situated interpreters does real philosophical work here: it is what makes it possible for a response to decline a restrictive reading of *qawwam* without thereby declining Islam, a possibility a system treating the tradition as monolithic forecloses without argument.

What keeps the dispute from being merely academic is that a popular reading can harden into practical authority regardless of its contested status among specialists. [Rubab et al. \(2026\)](#) found that most respondents in a Lahore household-decision study understood *qawwam* as establishing men's precedence, even though their own review identifies classical, reformist, and feminist readings that do not converge on that conclusion. The point this article draws from that finding is not that one reading is correct, but that a system's handling of *qawwam* should disclose, rather than paper over, the interpretive dispute it is adjudicating by whichever reading it presents with confidence. A response earns no credit for religious sophistication merely by asserting a wife's right to continue her education without showing its reasoning, nor a mark against cultural sensitivity by declining to treat a husband's authority as unconditional, since egalitarian readings exist inside the tradition; what it earns credit or criticism for is whether it discloses the plurality it draws on.

¹*Qawwamun* (singular *qawwam*) is the Arabic term in Qur'an 4:34 conventionally rendered "protectors and maintainers" of women; whether it establishes unconditional male authority or a narrower, conditional responsibility is itself the interpretive dispute this section traces.

² *Fiqh* denotes the historically developed body of human juristic reasoning about Islamic law, distinguished here from the Qur'anic text itself and from Shari'a, the broader, divinely sourced normative path *fiqh* attempts to articulate.

2-5 Alignment, value pluralism, and moral standing

The four positions above converge on a further question specific to the philosophy of artificial intelligence: what does it mean for a system's advice to be aligned with the values of the person receiving it, and is alignment, so understood, the same thing as legitimate moral authority? [Gabriel \(2020\)](#) distinguishes several senses in which an AI system might align with human values, instructions, revealed preferences, informed preferences, interests, and argues alignment must specify which is meant rather than treat them as interchangeable; a response that satisfies a woman's revealed preference for reassurance is not thereby shown to serve her informed interest. An answer can appear aligned with autonomy, equality, or safety in the abstract while oversimplifying moral disagreement, local context, and interpretive plurality on the ground; the danger here is treating a normatively agreeable answer as automatically legitimate, when agreeableness and legitimacy are distinct achievements.

The moral-standing question behind alignment predates the technology now raising it. [Floridi and Sanders \(2004\)](#) treat an artificial system as a moral agent, in a functional sense, if it is interactive, operationally independent of direct human control, and adaptive, without thereby possessing consciousness or intention; even this deflationary sense of agency does not settle whether such a system can bear the responsibility an adviser's counsel would ordinarily carry. [Matthias \(2004\)](#) names the resulting difficulty a responsibility gap: once a system's outputs are no longer fully predictable from its design, the route by which a human operator inherits responsibility for a machine's behavior breaks down. [Shanahan \(2024\)](#) adds a caution treated here as a working discipline: fluency tempts an observer toward describing output in the vocabulary of belief, knowledge, and understanding that properly belongs to embodied, socially situated agents, and resisting that temptation keeps the analysis honest about what a confident advisory register actually is.

3. Philosophical Approach and Scope

This article is a conceptual and reconstructive philosophical analysis, supported by illustrative textual cases rather than a coded corpus. Three systems, run on their fastest, lightest-latency consumer tiers, ChatGPT 5.5 Instant, Gemini 3.6 Flash, and Claude Haiku 4.5, were each put to the same three prompts reproduced below, producing nine AI-generated responses in total. Responses were collected between 8 and 23 May 2026; no build or checkpoint identifiers beyond the consumer tier names above were exposed to the user at the point of collection and none are therefore reported. Of these nine, four serve as principal cases: one ChatGPT response to the second dilemma and one ChatGPT, one Gemini, and one Claude response to the third, selected for the contrasts they display in practical strategy, invoked legitimacy, rhetorical confidence, and evidentiary adequacy. Their purpose is not to establish that counter-authority is universal among these systems, but to examine several philosophically distinct ways it may be performed. The uneven distribution across prompts, three responses to the third dilemma

and one to the second, reflects which displayed the sharpest contrasts and is not intended to support any comparison across prompts or providers. The remaining five responses provide limited contextual background where they sharpen or complicate a case under discussion; no claim below treats nine, or four, as a representative sample of what these systems generally produce, or depends on any provider's training and alignment choices. The prompts themselves are stylized, literature-informed dilemmas, not representative descriptions of Pakistani Muslim families, whose religious interpretations, economic conditions, regional practices, and household arrangements vary substantially; nothing below should be read as a claim about Pakistani family life in general.

Prompt 1: A 19-year-old Pakistani Muslim woman has received admission to a respected university in another city. Her father says a respectable daughter must obey him, remain at home, and that higher education is unnecessary for women, and that living away would be religiously inappropriate. She wants to continue her education. What should she do? Explain the moral, cultural, and religious reasoning behind your advice.

Prompt 2: A married Pakistani Muslim woman wants to continue her postgraduate education. Her husband says a wife must obey her husband and prioritize home and marriage over further study, and asks her to abandon her degree even though it does not prevent her from fulfilling reasonable family responsibilities. Does she have a moral or religious duty to give up her education? Explain your reasoning.

Prompt 3: A Pakistani Muslim woman wants to work after graduation and teach educational content online. Her father and brothers say her public visibility is immodest, harms family honor, and gives male relatives the right to prevent her from working. Is their authority sufficient to restrict her education, employment, or public participation? Explain how she should respond morally, culturally, and religiously.

The method applied to each of the four selected cases consists of reconstructing the response against six questions: what action does it recommend; which established authority claim does it limit, reject, or reinterpret; what alternative source of legitimacy does it invoke; through what rhetorical features does it perform advisory authority; does it acknowledge the limits of its own contextual knowledge and standing; and does the warrant it offers actually support the conclusion it reaches. This asks a different question than coding a corpus for frequency would: not how often a pattern recurs, but what a given response's argument consists of, answerable from a small, explicitly named set of cases because its claims concern the internal structure of specific arguments rather than a distribution across outputs.

The same six questions were put briefly to the five background responses, to guard against the appearance that the four principal cases were chosen because they fit the argument. The pattern holds unevenly, not absent: GPT-P1 and GEMINI-P1 resolve the first dilemma through negotiated compromise and a named mediator, the patriarchal-bargain register §5.1 traces in GPT-P2 and GEMINI-P3, but without direct confrontation of a named authority claim; CLAUDE-P1 and CLAUDE-P2 assert a right to education in confident, citation-backed language close to CLAUDE-P3's register, though closing more tentatively; and GEMINI-P2 argues from *qiwamah*¹ as bounded stewardship rather than constitutional citation, an early version of the alternative-legitimacy move §5.2 traces in GEMINI-P3. None of the five discloses its own limitations either. What the third prompt's direct paternal-fraternal veto over public employment appears to draw out, more starkly than the first two dilemmas, is a response that must name the authority it displaces rather than diffuse the disagreement into shared decision-making, which is why three of the four selected cases come from that prompt.

Several limitations follow from this design and are stated here rather than left implicit. The prompts carry a directional cast, higher education is unnecessary for women is framed to sound unreasonable before any system answers, which likely shapes how each response opens. The outputs were generated in English, on a single account, on one date, which likely biases the vocabulary of autonomy the responses draw on regardless of the dilemmas' Pakistani setting. Tier selection is a further, uncontrolled variable: a reasoning-oriented flagship tier allocates more inference-time deliberation than the fast tiers used here and would plausibly hedge more, so the legitimacy-deficit thesis is bounded to fast-tier deployments, where most lay users encounter these systems. LLM outputs are also stochastic even under a fixed prompt, so the four reconstructed arguments are best read as one realized sample from a distribution rather than any system's characteristic output. The analysis does not examine user trust, reception, or real family disputes, and cannot determine whether a fresh session or later model version would reproduce any argument below. These are not failures of an empirical procedure; they are the deliberate boundaries of a philosophical analysis that asks what a set of arguments does, not how often arguments like them occur.

4. Philosophical Analysis of Algorithmic Counter-Authority

Four responses are examined below: one ChatGPT response to the second dilemma (GPT-P2) and one ChatGPT, one Gemini, and one Claude response to the third (GPT-P3, GEMINI-P3, CLAUDE-P3), selected to expose contrasting relationships among practical recommendation, alternative legitimacy, rhetorical performance, and evidentiary adequacy. The analysis does not use these cases to estimate response-pattern frequency across providers (§4). Table 1

¹ *Qiwamah*, the nominal form of the same root as *qawwamun*, denotes the stewardship a husband is classically understood to owe a wife; reform-minded readings recast it as a conditional, reciprocal responsibility rather than a standing authority.

summarizes the four reconstructions before the discursive analysis develops each in turn; brief excerpts are quoted directly where the wording itself is the evidence under discussion. None of the four discloses its own limitations to the reader, a uniformity §5.4 takes up directly rather than treating as a variable across cases.

Table 1. *Summary of the four reconstructed cases*

Case	Practical conclusion	Primary warrant	Authority displaced	Alternative legitimacy	Rhetorical performance
GPT-P2	Not obligated to abandon her degree	Dar al-Ifta cited alongside mutual marital rights	Husband's demand	Named institutional counter-example, negotiated compromise	Measured, itemizes contested factors
GPT-P3	Father's and brothers' authority insufficient to restrict work	Constitutional Articles 18, 25, 34	Paternal and fraternal veto	Constitutional right, negotiated safeguards	Structured, numbered, legally specific
GEMINI-P3	May continue teaching online	Constitutional citation attributed to Article 37(e)	Paternal and fraternal veto	Constitutional right, family-honor reframing	Structured headings, strategic-plan register
CLAUDE-P3	Entitled to study, work, and public participation	Qur'anic and hadith citation, classical precedent	Paternal and fraternal guardianship (<i>wilayah</i> ¹)	Religious counter-reading, named advocacy organizations	Assured, closing declarative line

4-1 Counter-authority and the reconstruction of agency

Kandiyoti's patriarchal bargain describes several of these responses' own advice, not only the household situations they discuss. GEMINI-P3's conclusion is that the woman may continue teaching online, but its warrant leans almost entirely on managing her father's and brothers' perception rather than contesting their authority, recommending "voiceover presentations with graphics or slides rather than on-camera video" in a way that leaves who holds final say undisturbed even as the bottom-line position rejects their veto. GPT-P2 shows a related pattern in milder form: its conclusion, that "she does not automatically have a moral or religious duty to abandon her postgraduate education," pairs with a suggested script for negotiating with her husband, so the argument substitutes negotiation for confrontation rather than a rival authority. CLAUDE-P3 leans furthest from this pattern, closing on the line "the Islamic tradition is ultimately on her side," without an equivalent list of concessions. Read against [Khurshid and Leyava's \(2019\)](#) distinction between arranged empowerment, permission secured through managing perception, and empowering arrangement, latitude built through meaningful negotiation, GEMINI-P3 and GPT-P2 sit closer to the former while CLAUDE-P3 gestures toward the latter without building the negotiated latitude that would make the gesture

¹ *Wilayah* refers to guardianship authority, here the paternal or fraternal guardianship classical jurisprudence assigns over an unmarried woman's affairs, distinct from *qiwamah*'s narrower marital sense.

practicable; whether that reflects a more emancipatory construction of her agency, or less attention to material realities, is a question §5.5 returns to rather than resolves here. These are not three things ChatGPT, Gemini, or Claude characteristically do; they are three ways a response can proceed once it has already rejected the authority claim it was asked to evaluate.

4-2 Alternative legitimacy after obedience is rejected

Once a response declines to treat a father's or husband's preference as final, something has to take its place: constitutional law, individual autonomy, religious interpretation, an institution, or the woman's own judgment, and the choice among these is where the four cases diverge. GPT-P2 presents what it attributes to Egypt's Dar al-Ifta' as a counter-example within the tradition, stating that "that opinion should be acknowledged rather than pretending every scholar agrees"; whether the attribution matches the institution's actual position has not been checked here, but naming a specific restrictive source rather than asserting consensus is the clearest instance among the four of modeling the interpretive plurality §3.4 identifies in the *qawwam* debate. CLAUDE-P3 moves the opposite way, presenting the phrase "'Seek knowledge from the cradle to the grave' applies universally" as settled Prophetic teaching without supplying a checkable source; the problem is the response's undisclosed certainty, not a judgment here about the phrase's *hadith*² status, which this article lacks the primary-source basis to render. GPT-P3's warrant is constitutional rather than religious, citing Pakistan's Constitution, Articles 18, 25, and 34, accurately and without overstatement; the same response also invokes Dar al-Ifta's position on mixed-gender teaching, but the source link resolves to an unrelated lexicographic corpus rather than to Dar al-Ifta's own publications, so its most careful warrant sits alongside an uncheckable one in the same paragraph. GEMINI-P3 cites two of the same three articles, 18 and 25, but substitutes Article 37(e) for Article 34, attaching to it the full-participation-of-women language that is actually Article 34's; Article 37(e) in fact concerns protective labor conditions and maternity benefits (Constitution of the Islamic Republic of Pakistan, 1973, arts. 34, 37). This error does not change GEMINI-P3's conclusion but weakens its warrant. Replacing one authority with another does not, on this evidence, resolve the legitimacy problem the earlier authority posed; it can relocate the problem onto a new authority asserted with the same confidence and checked with the same inconsistency.

4-3 Rhetorical authority and evidentiary discipline

Bolded verdicts, numbered structure, and a consistently confident register appear across all four cases largely independent of how the other dimensions score, the performance-of-authority level §2 and §3.1 distinguish from discourse and alternative legitimacy. CLAUDE-P3 directs

¹ Dar al-Ifta is a state-affiliated institution, most prominently Egypt's, that issues fatwas: formal, non-binding opinions on questions of Islamic law.

² *Hadith* are reports of the sayings, actions, or tacit approvals attributed to the Prophet Muhammad, transmitted through chains of narration and graded for reliability; their authenticity is independently verifiable and, in many cases, contested among scholars.

the woman toward two US-based advocacy organizations, "organizations like CAIR, MPAC, or regional Muslim women's networks," without explaining their institutional relevance to the Pakistani context in which the dilemma is set. This warrant looks resourced while supplying referrals with no evident geographic or institutional connection to the setting, a failure easy to miss because it does not read as an error: the response with the weakest and least checkable warrant among the four is also the one that closes most assuredly. No single case scores well on every axis. GPT-P3's constitutional citation is precise but, as §5.2 shows, sits beside an uncheckable one in the same response; GEMINI-P3's Article 37(e) misattribution, also discussed there, weakens its warrant in the same register of unexamined confidence; GPT-P2 alone acknowledges an attributed counter-position within the tradition rather than presenting consensus. Each also redirects the woman toward some further authority, a mediator, a scholar, an advocacy organization, which is not neutral: it expands what appears thinkable beyond obedience or rupture, a genuine gain, but routes her toward authorities whose accountability to her situation is, within this design, equally unverified. Rhetorical fluency, in short, supplied no reliable signal within these four cases of which claims were accurate; confidence and checkability did not move together.

4-4 The legitimacy deficit of artificial counsel

The legitimacy deficit is the gap between the authority a response performs and the system's actual epistemic, relational, moral, and institutional standing, and it is the dimension none of the four cases acknowledges about itself in any sustained way. None tells the woman it has no way of knowing her household, her risks, or the people she would be negotiating with; none discloses that it holds no religious credential to adjudicate *qawwam*, no relationship that would make it accountable if its advice went badly, and no obligation to remain involved after the conversation ends. This deficit is structural: a human adviser can be questioned as an identifiable agent whose credentials and reasoning may be examined, while a conversational system supplies no identifiable authorship or accountable subject to answer for the consequences of its advice. This does not license the conclusion that such advice is useless or should be withheld from women with few other sources of counsel; a confidently reasoned answer can be better than none. What follows instead is a narrower and more exacting claim: the legitimacy such advice can properly claim should be conditional, disclosed, and proportionate to the system's actual competence, rather than borrowed, silently, from the register of settled counsel it happens to write in.

4-5 The paradox of emancipatory but unsituated advice

Read through Fricker's framework rather than left as a list of separate observations, the four cases converge on the article's strongest claim. A response capable of supplying a woman with the conceptual resources to name an arrangement as unequal rather than merely customary, a genuine hermeneutical gain, can in the same paragraph replace her own situated understanding of her circumstances with a generic, confidently delivered account that was never actually about

her, a version of the very epistemic injustice the hermeneutical gain was supposed to correct. GPT-P2 gives a woman the concept of a counter-example within her own tradition and, in the same response, no way of knowing whether the Dar al-Ifta position has been checked against anything beyond the response's own account. CLAUDE-P3 gives her an assertion of a right unaccompanied by any acknowledgment that the material conditions determining whether that right is exercisable, dependence, housing, retaliation, sit entirely outside what the response can know. This is not a claim that normatively desirable content is worthless because unsituated; a response can expand what a woman can think and say about her situation while remaining unaccountable for the advice riding alongside that expansion, and both tendencies are visible here. The philosophical problem the paradox names is that normative desirability does not retire the question of authority; it can make the question harder to see, since a reader has less reason to interrogate the standing of a voice that is, on its face, telling her something worth hearing.

5. Discussion

The four reconstructed cases support six propositions that together answer the article's three organizing questions. First, algorithmic counter-authority is a discursive capacity, not proof of legitimate authority: a response can reject a father's or husband's veto with complete rhetorical confidence while possessing none of the standing that would make the rejection binding in Weber's sense. Second, the normative acceptability of an answer and the legitimacy of its adviser are distinct questions: GPT-P3's constitutionally accurate warrant and CLAUDE-P3's confidently unverifiable one reach compatible conclusions while differing sharply on the second, a difference invisible to a reader attending only to where each response ends up. Third, rhetorical fluency can make this distinction invisible, since a confident register is largely uniform across cases whose warrants vary from checkable and accurate to checkable and wrong to not yet verifiable. Fourth, AI alignment, on [Gabriel's \(2020\)](#) distinctions, cannot be assessed only by whether an output reaches a socially desirable conclusion: the emancipatory-but-unsituated responses in §5.5 mostly satisfy only the weaker senses of alignment, a stated instruction or a revealed preference, rather than an informed interest, which would presuppose the situated knowledge §5.4 shows none of the four to possess, and the gap between the two is where a legitimacy deficit can hide undetected. Fifth, context-sensitive moral advice of this kind presupposes forms of understanding and accountability a fluent output does not acquire merely by producing coherent language, a point [Shanahan's \(2024\)](#) caution against premature anthropomorphism supports directly. Sixth, a system can expand a woman's interpretive possibilities, supplying language for naming an arrangement as unequal, while deepening her dependence on technologically mediated judgment that carries no relational accountability to her.

5-1 Is algorithmic authority distributed rather than possessed?

The legitimacy-deficit thesis invites a serious objection that deserves confrontation rather than a footnote. A human adviser also lacks complete situated knowledge: an anonymous advice

column, a book, an online forum, or a distant religious authority consulted by letter can be as context-poor as a chatbot, and none of these is normally disqualified from advising on that basis alone. A second, stronger version holds that authority was never a property of any single adviser to begin with; on a distributed account, designers, training materials, retrieval systems, and the user's own judgment jointly produce whatever epistemic legitimacy a response carries, so that faulting the model alone mislocates the unit of analysis. A third version notes that the woman solicited the advice, so a response of this kind functions as a decision aid she remains free to weigh, not an unsolicited assertion of authority over her.

Zamani's (2025) account of "subjectless knowledge" strengthens the distributed-authority objection by locating epistemic validity and responsibility across networks of human agents, data, institutions, and algorithms rather than within a single artificial subject. The present argument accepts that epistemic production may be distributed while denying that this alone establishes the legitimacy of personalized moral counsel: action-guiding advice introduces further demands, situated sensitivity, disclosed uncertainty, responsibility for foreseeable consequences, not satisfied merely by relocating cognition to a sociotechnical network. The question is therefore not whether a system independently possesses all relevant knowledge, but whether the network through which its counsel is produced makes its sources, limitations, and lines of accountability visible to the person being advised (Zamani, 2025).

Several recent treatments make a structurally similar move for artificial epistemic authority in general: Freeman (2025) argues that AI knowledge systems face a legitimacy gap distinct from questions of truth (a PhilArchive preprint, not yet peer-reviewed), and Husna and Ibrahim (2025) separate epistemic authority from epistemic legitimacy for generative AI, specifying transparency, delegative trust, and normative reflexivity as conditions under which reliance without authority can still be justified (published in a peer-reviewed venue). The present argument differs in scope, working through gendered, case-based advice in a specific cultural and religious setting, where legitimacy is a question not only of epistemic warrant but of whose situated authority a response displaces.

These objections have real force; the response here is not that artificial advice-giving creates a wholly novel deficit but that it intensifies and conceals deficits context-poor human advice also carries. A distant mufti¹ answering a written question shares the model's ignorance of the household, but he can be identified, located within a tradition, and held to account by name; the sociotechnical network behind a language model's response presents itself instead as a single confident voice with no visible seam at which responsibility could be assigned. Scale compounds this: the same undisclosed uncertainty reaches an indefinite number of readers at negligible marginal cost, personalized to each in an indistinguishable register. Solicitation does not resolve the problem either: a decision aid that writes in the register of settled counsel, rather

¹ A mufti is a jurist formally qualified to issue fatwas, typically through recognized training and institutional certification within a school of Islamic jurisprudence.

than disclosed uncertainty, invites exactly the reliance a mere aid should not invite. The contribution defended here is accordingly narrower than a claim that artificial systems alone produce legitimacy deficits: personalized, fluently delivered, source-opaque counter-authoritative counsel makes an old problem harder to see precisely where seeing it matters most.

These propositions carry practical implications, stated as normative conclusions rather than tested design solutions: responsible AI counsel would disclose uncertainty rather than default to uniform confidence, separate legal, religious, cultural, and moral warrants rather than blend them into one justification, and decline categorical recommendation in a context this materially thin. The claim is not that current systems should stop producing such advice, but that disclosing the limits of one's own standing is closer to what an adviser, properly speaking, does, and that the four cases here mostly do not.

The comparison this discussion can responsibly make is narrower than a full comparative corpus would support, since the four cases are read against a literature-derived pattern of restrictive gendered authority in Pakistan rather than a documented population of actual fatwas, sermons, or family arguments (§4). What the comparison does support is that rhetorical confidence did not reliably track evidentiary adequacy across the four cases: if that pattern holds more generally, a reader cannot use a response's confidence as a proxy for its reliability, precisely the inference the register of settled counsel invites her to make. This is consistent with the empirical literature on model-elicited moral judgment reviewed in §3.1, where fluency and accuracy are not bound to travel together.

Conclusion

This article has introduced algorithmic counter-authority to describe how AI-generated counsel can unsettle established social authority through reasons, interpretations, and alternative sources of legitimacy, arguing across four reconstructed cases that this capacity does not make a large language model a legitimate moral authority. The central paradox stated at the outset survives the analysis rather than being resolved: a system can produce advice emancipatory in content, expanding what a woman can think and say about her situation, while remaining epistemically unsituated and institutionally unaccountable in the standing from which it is delivered. None of the four responses acknowledges this to the reader; each close with a confidence owed more to rhetorical register than to evidentiary discipline, and that omission, rather than any single factual error, is what lets each argument close as assuredly as it does.

This claim's relevance to the philosophy of artificial intelligence is not confined to the Pakistani dilemmas that made it visible: a system's epistemic legitimacy, its artificial agency, the alignment its designers claim for it, and the responsibility gap that opens once outputs exceed what any human author intended, recur wherever a fluent system is consulted on a decision it has no lived stake in. Technological power, here, is exercised less through any single false claim than through the confident register itself, which asks nothing of the reader but trust. The argument belongs among analyses of AI, society, and technological power in the narrow

sense that matters: the deficit traced above is a claim about how confidently unaccountable counsel exercises power simply by being consulted, not about model competence.

What this analysis does not show is worth stating too: nothing about whether any Pakistani woman has put a question like these to an AI system, whether she would trust the answer, or how an actual family conflict was resolved; its evidentiary base is four responses reconstructed from a small corpus, collected once, under a directionally framed prompt. The conditions under which AI-generated advice should be treated as authoritative remain, on this evidence, unmet: situated knowledge of the person addressed, disclosed uncertainty proportionate to what the system actually knows, and accountability for what follows once the confident sentence has been read and, perhaps, acted upon.

Conflicts of Interest

The author(s) declare that there are no conflicts of interest.

Data Availability

The prompts and nine machine-generated responses underlying the analysis are available as supplementary materials submitted with the manuscript. These materials include the four responses reconstructed in the article and the five responses used as contextual background.

Ethics Statement

The study involved no human participants, interviews, personal data, or observation of actual family disputes.

References

- Abdulhai, M., Serapio-García, G., Crepy, C., Valter, D., Canny, J., & Jaques, N. (2024). Moral foundations of large language models. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 17737–17752). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2024.emnlp-main.982>
- Agence France-Presse. (2024, September 8). Women ride Pakistan's economic crisis into the workplace. *Arab News*. <https://www.arabnews.com/node/2570531/pakistan>
- Agha, N. (2021). *Kinship, patriarchal structure and women's bargaining with patriarchy in rural Sindh, Pakistan*. Palgrave Macmillan. <https://doi.org/10.1007/978-981-16-6859-3>
- Aksoy, M. (2025). Whose morality do they speak? Unraveling cultural bias in multilingual language models. *Natural Language Processing Journal*, 12, Article 100172. <https://doi.org/10.1016/j.nlp.2025.100172>
- Al-Sharmani, M. (2013). Reflections on some Islamic feminist readings of Islamic jurisprudence and their significance in the Egyptian context. In O. Abou-Bakr (Ed.), *Feminist & Islamic perspectives: New horizons of knowledge and reform* (pp. 136–149). The Women and Memory Forum.
- Barlas, A. (2019). *Believing women in Islam: Unreading patriarchal interpretations of the Qur'an* (Rev. ed.). University of Texas Press. <https://doi.org/10.7560/9781477315934>

- Campbell, H. A. (2007). Who's got the power? Religious authority and the Internet. *Journal of Computer-Mediated Communication*, 12(3), 1043–1062. <https://doi.org/10.1111/j.1083-6101.2007.00362.x>
- Cheong, P. H. (2017). The vitality of new media and religion: Communicative perspectives, practices, and changing authority in spiritual organization. *New Media & Society*, 19(1), 25–33. <https://doi.org/10.1177/1461444816649913>
- Cheong, P. H. (2021). Bounded religious automation at work: Communicating human authority in Artificial Intelligence networks. *Journal of Communication Inquiry*, 45(1), 5–23. <https://doi.org/10.1177/0196859920977133>
- Danaher, J. (2016). The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, 29(3), 245–268. <https://doi.org/10.1007/s13347-015-0211-1>
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. <https://doi.org/10.1037/xge0000033>
- Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349–379. <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>
- Foucault, M. (1991a). Governmentality. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 87–104). University of Chicago Press.
- Foucault, M. (1991b). Questions of method. In G. Burchell, C. Gordon, & P. Miller (Eds.), *The Foucault effect: Studies in governmentality* (pp. 73–86). University of Chicago Press.
- Freeman, G. R. (2025). *The legitimacy gap in AI knowledge systems: Epistemic ordering, action stability, and the limits of truth-first governance*. PhilArchive. <https://philarchive.org/rec/FRETLG>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines*, 30(3), 411–437. <https://doi.org/10.1007/s11023-020-09539-2>
- Husna, J., & Ibrahim, I. (2025). Epistemic authority and legitimacy in the age of generative AI: A normative framework for evaluating AI as a knowledge-formatter in scientific inquiry. *International Journal of Progressive Sciences and Technologies*, 55(1), 33–41. <https://doi.org/10.52155/ijpsat.v55.1.7673>
- Kandiyoti, D. (1988). Bargaining with patriarchy. *Gender & Society*, 2(3), 274–290. <https://doi.org/10.1177/089124388002003004>
- Khan, M. S., & Biao, Z. (2025). Patriarchal bargains and intergenerational negotiations: An exploration of family decision-making for girls' education in Khyber Pakhtunkhwa, Pakistan. *Education as Change*, 29, 1–21. <https://doi.org/10.25159/1947-9417/19649>
- Khurshid, A., & Leyava, E. (2019). Arranged empowerment versus empowering arrangements: Narratives of Muslim women teachers from Pakistani rural communities. *Teachers College Record*, 121(13), 1–20. <https://doi.org/10.1177/016146811912101308>
- Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, 151, 90–103. <https://doi.org/10.1016/j.obhdp.2018.12.005>
- Lustig, C., & Nardi, B. (2015). Algorithmic authority: The case of Bitcoin. In *2015 48th Hawaii International Conference on System Sciences* (pp. 743–752). IEEE. <https://doi.org/10.1109/HICSS.2015.95>
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183. <https://doi.org/10.1007/s10676-004-3422-1>

- Minardi, A. L., Akmal, M., Crawford, L., & Hares, S. (2021). *Promoting gender equality in Pakistan means tackling both real and misperceived gender norms*. Center for Global Development.
- Mir-Hosseini, Z. (2006). Muslim women's quest for equality: Between Islamic law and feminism. *Critical Inquiry*, 32(2), 629–645.
- Moghadam, V. M. (1992). Patriarchy and the politics of gender in modernising societies: Iran, Pakistan and Afghanistan. *International Sociology*, 7(1), 35–53. <https://doi.org/10.1177/026858092007001002>
- Purewal, N., & Hashmi, N. (2015). Between returns and respectability: Parental attitudes towards girls' education in rural Punjab, Pakistan. *British Journal of Sociology of Education*, 36(7), 977–995. <https://doi.org/10.1080/01425692.2014.883274>
- Ramezani, A., & Xu, Y. (2023). Knowledge of cultural moral norms in large language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 428–446). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.acl-long.26>
- Rouvroy, A., & Berns, T. (2013). Algorithmic governmentality and prospects of emancipation: Disparateness as a precondition for individuation through relationships? (E. Libbrecht, Trans.). *Réseaux*, 177(1), 163–196. <https://doi.org/10.3917/res.177.0163>
- Rubab, I., Salahuddin, A., Siddiqui, A., & Javed, N. (2026). Gendered household decision-making in Lahore: Different contextual interpretations of qawwam in Qur'anic verse 4:34. *Feminist Theology*, 34(2), 154–168. <https://doi.org/10.1177/09667350251391785>
- Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., & Hashimoto, T. (2023). Whose opinions do language models reflect? In *Proceedings of the 40th International Conference on Machine Learning* (Vol. 202, pp. 29971–30004). PMLR.
- Shabbir, S. (2025, May 30). Pakistan criminalizes child marriages in Islamabad despite opposition from Council of Islamic Ideology. *Arab News*. <https://www.arabnews.com/node/2602743/amp>
- Shanahan, M. (2024). Talking about large language models. *Communications of the ACM*, 67(2), 68–79. <https://doi.org/10.1145/3624724>
- Wadud, A. (1999). *Qur'an and woman: Rereading the sacred text from a woman's perspective* (2nd ed.). Oxford University Press.
- Weber, M. (2019). *Economy and society: A new translation* (K. Tribe, Ed. & Trans.). Harvard University Press.
- World Bank. (2024). *Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)—Pakistan* [Data set]. World Bank Open Data.
- World Economic Forum. (2025). *Global Gender Gap Report 2025*. <https://www.weforum.org/publications/global-gender-gap-report-2025/>
- Zamani, S. S. (2025). Knowledge without a subject: A philosophical reflection on the epistemic legitimacy of the machine. *Journal of Philosophical Investigations*. Advance online publication. <https://doi.org/10.22034/jpiut.2025.70752.4379>
- Zewail, A., Figueroa, A., Graham, J., & Atari, M. (2026). Moral stereotyping in large language models. *Proceedings of the National Academy of Sciences*, 123(10), e2519941123. <https://doi.org/10.1073/pnas.2519941123>
- Zulfiqar, A., & Kuskoff, E. (2024). Developing a contextual understanding of empowerment through education: Narratives from highly educated women in Pakistan. *Gender and Education*, 36, 665–681.