

# Intelligent Machines, Ethics and Law (COMP2400/6400)



## AI and Epistemic Injustice – Lecture

**Dr Regina Fabry**

Discipline of Philosophy, School of Humanities

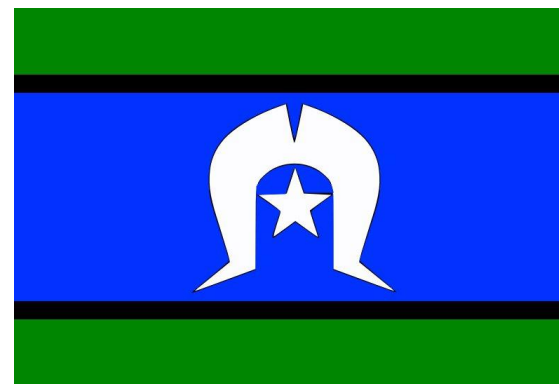
Macquarie University

Week 3 – 10 March 2025



**MACQUARIE**  
University  
SYDNEY • AUSTRALIA

*We acknowledge the Traditional Custodians of the land on which Macquarie University stands – the Wallumattagal Clan of the Dharug Nation – whose cultures and customs have nurtured, and continue to nurture, this land since time immemorial. We pay our respects to the Elders, past and present.*

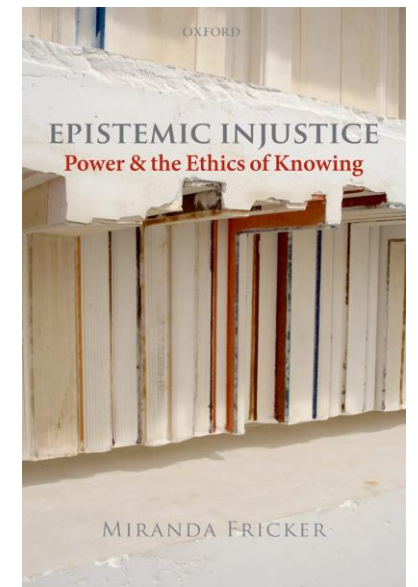


## Introduction: 5 Questions

1. What is epistemic injustice?
2. What is testimonial injustice?
3. What is hermeneutical injustice?
4. How can algorithmic profiling facilitate or exacerbate forms of epistemic injustice?
5. How can generative AI systems facilitate or exacerbate forms of epistemic injustice?

# What Is Epistemic Injustice?

- The notion of ‘epistemic injustice’ refers to “[...] a wrong done to someone specifically in their capacity as a knower.” (Fricker, 2007, p. 1)
- We exercise our epistemic capacities through various socially shaped practices, including conceptualisation, testimony, and meaning-making.
- If our epistemic capacities are impeded, truncated, or undermined, owing to an *identity prejudice* (Fricker, 2007, p. 4) or structural oppression (Young, 1990), we become targets of epistemic injustice.



# What Is Epistemic Injustice?

## Testimonial Injustice

- Testimonial injustice typically occurs in conversational exchanges between a speaker (or writer) and a hearer (or reader).
- A speaker suffers “[...] a testimonial injustice if and only if she received a credibility deficit owing to identity prejudice in the hearer; so the central case of testimonial injustice is *identity-prejudicial credibility deficit*.” (Fricker, 2007, p. 28; italics in original)
- A speaker’s testimony receives less credibility owing to the application of an (implicit, yet systematic) negative stereotype against their perceived or assumed (intersectional) social identity.

# What Is Epistemic Injustice?

## The Moral Harmful Wrong of Testimonial Injustice

The moral harmful wrong of testimonial injustice has two aspects:

1. The targets “[...] are degraded qua knower, and they are symbolically degraded qua human.” (primary harm) (Fricker, 2007, p. 44)
2. As a consequence of the double degradation, targets of testimonial harms suffer practical and epistemic disadvantages (secondary harm) (see Ibid., pp. 46-48):
  - a. Disadvantages regarding one’s legal or professional standing
  - b. Diminishment or loss of a sense of epistemic confidence and a sense of epistemic agency.

## What Is Epistemic Injustice? Hermeneutical Injustice

- Systematic, as opposed to incidental, cases of hermeneutical injustice are defined as “[...] the injustice of having some significant area of one’s social experience obscured from collective understanding owing to a structural identity prejudice in the collective hermeneutical resource.” (Fricker 2007, p. 155; italics removed)
- The notion of ‘structural identity prejudice’ captures a systematic, often implicit prejudice owing to the (often intersectional) social identity of members of structurally oppressed groups.
- ‘Collective hermeneutical resource’ is an umbrella term for widely shared concepts, socio-cultural patterns, or narratives for knowledge- and meaning-making.

# What Is Epistemic Injustice?

## The Sources of Hermeneutical Injustice

- According to Medina (2017), hermeneutical injustice can have (at least sometimes co-occurring) sources.
- Depending on the source, we can distinguish two different kinds:
- **Performatively** produced hermeneutical injustice: Agents are not treated as knowers “because of their communicative performance or expressive style” (Medina, 2017, p. 46).
- **Semantically** produced hermeneutical injustice: Concepts or labels for capturing socially shaped experiences are either not available or socially sanctioned.



José Medina



# What Is Epistemic Injustice?

## The Moral Harmful Wrong of Hermeneutical Injustice

- The moral harmful wrong of hermeneutical injustice can be characterised as *situated hermeneutical inequality* (Fricker, 2007, p. 162).
- It is defined as follows: “[...] the concrete situation is such that that subject is rendered unable to make communicatively intelligible something which it is particularly in his or her interests to be able to render intelligible” (Ibid.).
- Hermeneutical injustice prevents epistemic agents with membership in (intersecting) structurally oppressed groups to exercise their epistemic agency.
- It deprives structurally oppressed epistemic agents of significant resources for generating and sharing knowledge and engaging in meaning-making practices.

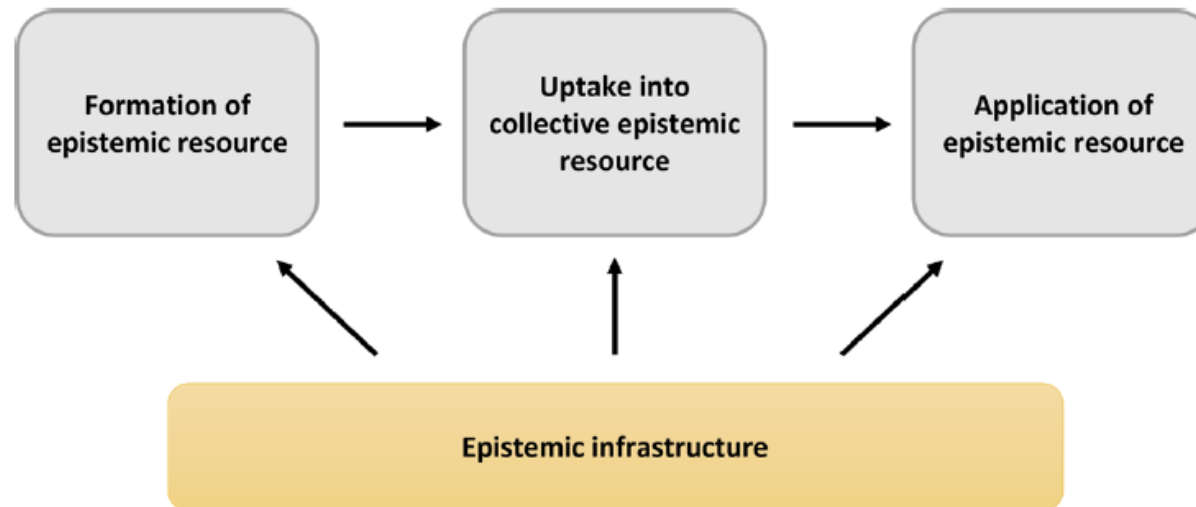


# **Hermeneutical Injustice and Algorithmic Profiling**

# What Is Algorithmic Profiling?

- The notion of ‘algorithmic profiling’ captures “[...] the automated process of extrapolating information about a person based on personal data.” (Milano & Prunkl, 2025, p. 191)
- Machine learning algorithms are deployed to predict preferences or needs across a wide range of contexts (e.g., consumer behaviour, interests in politics, popular culture, music, or sports).
- In online environments, algorithmic profiling is widely used across social media, search engines, and news websites.
- “[...] algorithmic profiling is used to determine which posts, search results, products, or news a given user sees at a given time.” (Ibid., p. 191)

# Situating Knowledge in Epistemic Infrastructures



**Fig.1** The diagram illustrates the different steps involved in the formation, collective uptaking, and application of epistemic resources. Sources of epistemic injustice are linked to dysfunctions within the various steps. Underlying the entire process is the epistemic infrastructure that creates opportunity for the sharing and comparing of experiences, thereby enabling epistemic resources to be created, taken up, and applied

# Hermeneutical Injustice Revisited

- In the context of algorithmic profiling, Milano & Prunkl (2025) propose, the relevant category of epistemic injustice is semantic hermeneutical injustice as conceptualised by Medina (2017).
- The availability of epistemic resources (e.g., concepts) can be dysfunctional at the following stages (see Milano & Prunkl, 2025, pp. 188-191):
  1. **Formation:** The engineering of new concepts or the development of labels and patterns for knowledge generation and meaning-making are disrupted.
  2. **Uptake:** Newly formed epistemic resources are silenced or socially sanctioned.
  3. **Application:** Newly formed and shared epistemic resources are incorrectly applied within and across contexts.

# Epistemic Problems of Algorithmic Profiling

## The Problem of Inference

- Algorithmic profiling can be associated with “[...] various shortcomings related to the extrapolation of information from collected datasets that might result in flawed, wrong, or inadequate profiles.” (Milano & Prunkl, 2025, p. 191)
- In many cases, this problem might be rooted in two other problems: the *proxy problem* (Johnson, 2021) and the *ground lies problem* (Bender, 2024) (→ W2: Algorithmic Bias).



# Epistemic Problems of Algorithmic Profiling

## The Problem of Inquiry

- Algorithmic profiling affects our epistemic self-determination while being epistemically opaque.
- Targets of algorithmic profiling have not access to their profiles, and the ML algorithms generating them, for the following (often co-occurring) reasons:
  - a) “individuals are not aware that they are subject to profiling”
  - b) “they are aware that they are subject to profiling but cannot access their profiles due to technical or institutional hurdles”
  - c) “they can access their profiles but cannot meaningfully interpret their profiles or the inferences that are performed on their basis: (Milano & Prunkl, 2025, p. 192)”

# Epistemic Fragmentation

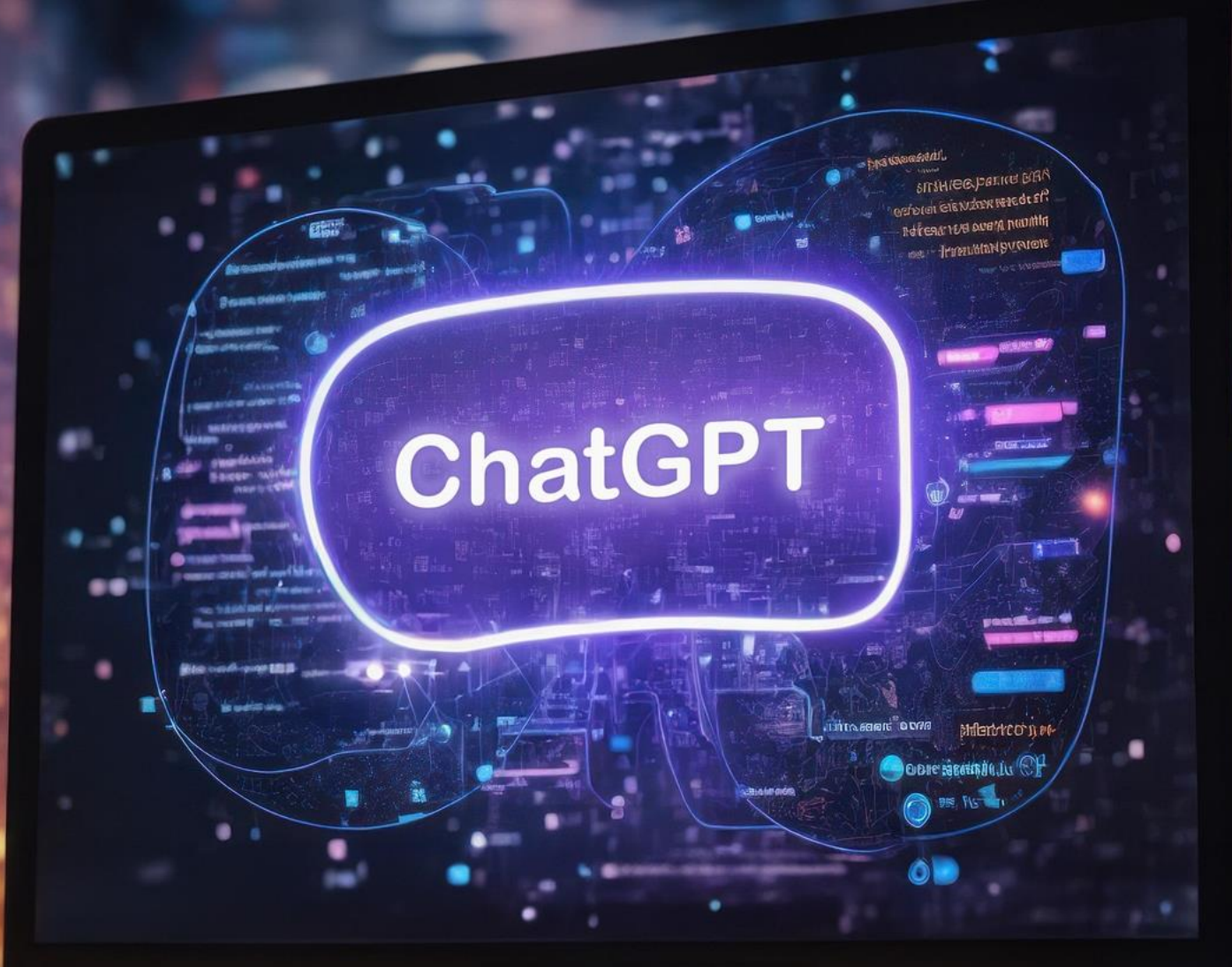
- Given the *problem of inference* and the *problem of inquiry*, algorithmic profiling often leads to epistemic fragmentation.
- Epistemic fragmentation can be defined as follows: It is a “[...] state in which individual epistemic agents have no (or severely limited) access to information about other individuals’ personal contexts.” (Milano & Prunkl, 2025, p. 192)
- Due to epistemic fragmentation, agents have insufficient epistemic resources for knowledge- and meaning-making.



# Epistemic Fragmentation and Hermeneutical Injustice

- Epistemic fragmentation can lead to hermeneutical injustice, because it causes a dysfunction at the following stages (see Milano & Prunkl, 2025, p. 193):
  1. **Formation:** It prevents meaningful social exchanges that could lead to epistemic engineering.
  2. **Uptake:** It prevents the distribution and sharing of epistemic resources.
  3. **Application:** It prevents the wide-spread adequate application of epistemic resources.





## Generative AI and Epistemic Injustice



# Generative Epistemic Injustice

- According to Kay et al. (2024), generative AI systems, including chatbots based on Large Language Models (e.g., ChatGPT) and text-to-image generators (DALL-E), enable and facilitate the rise of a new kind of epistemic injustice: ‘generative algorithmic epistemic injustice.’

1. **Generative amplified testimonial injustice:** when generative AI magnifies and produces socially biased viewpoints from its training data.
2. **Generative manipulative testimonial injustice:** when humans fabricate testimonial injustices with generative AI.
3. **Generative hermeneutical ignorance:** when generative AI lacks the interpretive frameworks to understand human experiences.
4. **Generative hermeneutical access injustice:** when unequal access to information and knowledge is facilitated by generative AI.

Kay et al. (2024), p. 687.

## Generative Amplified Testimonial Injustice

- A speaker suffers “[...] a testimonial injustice if and only if she received a credibility deficit owing to identity prejudice in the hearer; so the central case of testimonial injustice is *identity-prejudicial credibility deficit*.” (Fricker, 2007, p. 28; italics in original)
- Given training data that are biased or non-representative of structurally oppressed groups (→ W2: Algorithmic Bias), generative AI systems often perpetuate and amplified already existing forms of testimonial injustice.
- “In the algorithmic setting, the injustice requires a credibility excess assigned to the algorithm; that is, humans believe the account amplified through the technology over the individual or group who is discredited.” (Kay et al., 2024, p. 687)

# Generative Manipulative Testimonial Injustice

- In contrast to Fricker (2007), who holds that testimonial injustice often proceeds unintentionally, Kay et al. (2024) suggest that generative testimonial injustice can be a result of intentional and wilful manipulation.
- Manipulative testimonial injustice “[...] occurs when humans intentionally steer the AI to fabricate falsehoods, discrediting individuals or marginalised groups.” (Kay et al., 2024, p. 688)
- Deepfakes and other AI outputs are exacerbating the testimonial injustices suffered by epistemic agents that are members of structurally oppressed groups.

# Generative Manipulative Testimonial Injustice

- In contrast to Fricker (2007), who holds that testimonial injustice often proceeds unintentionally, Kay et al. (2024) suggest that generative testimonial injustice can be a result of intentional and wilful manipulation.
- Manipulative testimonial injustice “[...] occurs when humans intentionally steer the AI to fabricate falsehoods, discrediting individuals or marginalised groups.” (Kay et al., 2024, p. 688)
- Deepfakes and other AI outputs are exacerbating the testimonial injustices suffered by epistemic agents that are members of structurally oppressed groups.

# Generative Wilful Hermeneutical Ignorance

- According to Pohlhaus (2012, p. 416), “[...] wilful hermeneutical ignorance describes instances where marginally situated knowers actively resist epistemic domination through interaction with other resistant knowers, while dominantly situated knowers nonetheless continue to misunderstand and misinterpret the world.”
- In the context of generative AI, the notion of wilful hermeneutical ignorance captures “[...] how these systems can erase or misportray marginalized groups due to a lack of contextual and cultural understanding.” (Kay et al. 2024, p. 689).
- In contrast to other forms of wilful hermeneutical ignorance, “[...] generative hermeneutical ignorance is unique in that it stems directly from the limitations of generative AI models themselves.” (Ibid.)

## Generative Hermeneutical Access Injustice

- The harmful wrong of hermeneutical injustice partly lies in the inaccessibility of critical epistemic resources for members of structurally oppressed groups.
- In the context of generative AI, “[...] it centers on the generative AI’s control over access to information, leading to a denial of knowledge based on identity-driven bias or misrecognition.” (Kay et al. 2024, p. 689)
- Generative AI systems are predominantly trained with data stemming from WEIRD populations (Henrich et al., 2010) that are predominantly white and Anglophone (Bender, 2024).
- Accordingly, AI generated outputs are deprived of epistemic resources that are representative of and epistemically beneficial for members of structurally oppressed groups.



## Generative Algorithmic Epistemic Injustice

1. **Generative amplified testimonial injustice**: when generative AI magnifies and produces socially biased viewpoints from its training data.
2. **Generative manipulative testimonial injustice**: when humans fabricate testimonial injustices with generative AI.
3. **Generative hermeneutical ignorance**: when generative AI lacks the interpretive frameworks to understand human experiences.
4. **Generative hermeneutical access injustice**: when unequal access to information and knowledge is facilitated by generative AI.

## Concluding Summary

- The notion of ‘epistemic injustice’ refers to “[...] a wrong done to someone specifically in their capacity as a knower.” (Fricker, 2007, p. 1)
- Epistemic injustice can be testimonial by harmfully wronging the credibility of an agent’s testimony.
- It can be hermeneutical by harmfully wronging an agent through a lack or inaccessibility of epistemic resources.
- Epistemic injustices are facilitated, reinforced, or exacerbated through contemporary AI systems, e.g., algorithmic profiling (Milano & Prunkl, 2025) and generative AI (Kay et al. 2024).


# And Next...

## ... Blame and Responsibility in AI Systems

Philosophy & Technology (2021) 34:1057–1084  
<https://doi.org/10.1007/s13347-021-00450-x>

### RESEARCH ARTICLE



## Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them

Filippo Santoni de Sio<sup>1</sup>  · Giulio Mecacci<sup>2</sup>

Synthese (2023) 201:21  
<https://doi.org/10.1007/s11229-022-04001-5>

### ORIGINAL RESEARCH

## The risks of autonomous machines: from responsibility gaps to control gaps

Frank Hindriks<sup>1</sup>  · Herman Veluwenkamp<sup>1,2</sup> 

## References

- Bender, E. M. (2024). Resisting dehumanization in the age of “AI.” *Current Directions in Psychological Science*, 33(2), 114–120. <https://doi.org/10.1177/09637214231217286>
- Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The WEIRD people in the world? *The Behavioral and Brain Sciences*, 33(2–3), 61–135. <https://doi.org/10.1017/S0140525X0999152X>
- Johnson, G. M. (2021). Algorithmic bias: On the implicit biases of social technology. *Synthese*, 198(10), 9941–9961. <https://doi.org/10.1007/s11229-020-02696-y>
- Kay, J., Kasirzadeh, A., & Mohamed, S. (2024). Epistemic injustice in generative AI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 7, 684–697. <https://doi.org/10.1609/aies.v7i1.31671>
- Medina, J. (2017). Varieties of Hermeneutical Injustice. In I. J. Kidd, J. Medina, & G. Pohlhaus (Eds.), *The Routledge Handbook of Epistemic Injustice* (pp. 41–52). Routledge.
- Milano, S., & Prunkl, C. (2025). Algorithmic profiling as a source of hermeneutical injustice. *Philosophical Studies*, 182(1), 185–203. <https://doi.org/10.1007/s11098-023-02095-2>
- Pohlhaus Jr., G. (2012). Relational knowing and epistemic injustice: Toward a theory of willful hermeneutical ignorance. *Hypatia*, 27(4), 715–735. <https://doi.org/10.1111/j.1527-2001.2011.01222.x>
- Young, I. M. (1990). Five faces of oppression. In *Justice and the politics of difference* (pp. 39–65). Princeton University Press.