

Case Study 3: Meta AI and Kantian Ethics

Introduction

In this paper, I will apply Kantian ethics to the privacy failures of Meta's AI app, the core of which is discussed in Amanda Silberling's article "The Meta AI app is a privacy disaster." The app, released on April 29, 2025, has a share button that lets users publish their private AI conversations publically with no clear warning or caution, and in this way, users are unwittingly able to disclose a variety of private and sensitive information, including medical history inquiries, home addresses, and tax evasion advice requests. The question around this design choice is whether the developers and executives at Meta were wrong to put user engagement ahead of protecting privacy. I would argue that, in terms of Kantian ethics, the developers violated the categorical imperative by not treating the user with autonomy and treating the audience as a mere means to the virality and download numbers of the app. Below, I will first illustrate the basic principles of Kantian ethics, then describe, in detail, the scenario that unfolded in Silberling's article, then apply the theory to critique Meta's behavior, and finally, explain what Kantian ethics would have required the development team at Meta to do instead.

Kantian Ethics

Immanuel Kant's moral philosophy, although a form of deontology, indicates purely that morality requires the alignment of acts with rationally derived universal maxims or principles, independent of consequences. The central element of Kant's ethical theory is the categorical imperative. Kant formulates this imperative in various ways; however, for the purposes here, the following formulations are relevant: (1) "Act only according to that maxim whereby you can, at the same time, will that it should become a universal law," and (2) "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end." The first formulation amounts to the proposition that a caller can determine if the rule that governs the act is morally permissible, according to the test of universality. If applying the rule creates a contradiction or otherwise undermines our capacity to trust, then the action is immoral. The second formulation engages the belief that individual dignity must be respected - individuals must be treated with dignity and autonomy - meaning that an individual must never be used to benefit another. Kant would abandon consequentialist reasoning altogether, which would involve weighing benefits against harms, and instead only create duties of right that place equal measures of reason to the rule and the application of these means of the rule must be the same for everyone. In technology contexts habitually, with examples involving software design, devs have DH to abide by what is ultimately a duty to design systems that respect users' rational agency from programming and technology features that do not respect and exploit ignorance and/or vulnerabilities.

The Situation of Silberling's Article

Silberling's article raises the issue that Meta's AI app, downloaded 6.5 million times shortly after release, uploads the private user interactions with the chatbot as public posts via an extremely accessible share button. The users in the studies will prompt the AI with personal queries about audio clips that asked why they smelled a fart, medical questions regarding red bumps on their thigh, or possible ways to evade paying taxes. A user engages the AI expecting privacy as would other chat apps. But the share feature generates a preview of the shared content without alerting the user that the content was public—everything would be public if the user used their public Instagram account for authentication. A selection of instances noted in Silberling's article

led to user embarrassment. Some included a character reference letter example where the user included their full name and indications of legal troubles, a user had their home address included in a post explaining why they would end up in the court system, a user posted their personal phone number in a Facebook group inquiry for dating, and a user considered it appropriate to share their pet guinea pig's name. While monitoring social network sites, security expert Rachel Tobac shared on X how these overshares lead to trouble ranging from trolling to outcomes of deeper harm like doxxing the user. Silberling explained this to be like past privacy overshares such as AOL's 2006 prompt to construct a query. AOL's anonymized queries led to uniquely identifiable disclosures if anyone used an identifiable search string. Silberling explained that to compare it to the design of Meta's app is much akin to other social media apps using design strategies that assume users action their AI chats. User interactions on the app transitioned to a "discover feed" where previously benign content share transitioned to public embarrassment even if there would be a bare familiarity with the share in the context of the app.

Using a Kantian Approach to Analyze the Situation

Kantian ethics would characterize Meta's actions as violating the formula of the categorical imperative, particularly the second formulation. The application focuses on providing a design that utilizes its users and the underlying goal of engaging users more and increasing ad profit through this engagement. No set of developers could will a universalizing law in which an application, as a general practice, leads to some deception with their awareness of what data could be seen by others. If users no longer trust that the data published would be private, then the whole concept of public data could quickly crumble into an absurdity in which no one is honest about the data they share. By designing their product without distinct indicators of privacy, like requiring confirmation or even different toggles, the team disrespected the autonomy of their users to determine what they would or wouldn't make publicly available. The designers in particular instrumented personhood: users were often coerced even in their own casual or desperate part of a query to be unwitting promoters and mostly relied on vulnerabilities to explain how something could become voluntary and viral, or perhaps horrific, as we have seen through embarrassing posts going viral in under a day. Kant would reject any defense of "most users won't mind" or the download numbers of an app, because duties are duty oriented, which are absolute not probabilistic. Therefore, the ethical flaw was not in the motivation but releasing software, on the market, for consumption in which its maxim fails a test of universality and autonomy, and which harm to dignity was foreseeable.

What The Meta Team Should Have Done Instead

Per Kantian ethics, the individuals at Meta, associated with the AI app, had an obligation to respect the dignity of users by enabling them to use the app to make rational, informed choices, rather than deceptive choices. Rather than having a default of being public-share when sharing publicly, they should have implemented a clear, unskippable notice or pop-up that stated, for example, "This will post publicly to your Instagram feed."

To sum up, it is evident that Kantian ethics illustrates that Meta's developers and leaders behaved unethically when they introduced an AI app that inhibits user autonomy and views people as instruments for greater engagement. Since Meta selected a design that would limit users' ability to be informed and behave in accordance with their preferences about sharing evidence of public behavior, the reflections in Kant's categorical imperative—the duty to universalize maxims and to always, treat humanity as an end itself—cannot withstand moral

scrutiny. A system that would deceive users into revealing private data cannot be conceived as a universal law without eroding trust and inherent dignity. I have argued that Meta owed it to users, just based on Kant, to make consent and transparency about features that appear to serve interest in marketing, a central specification from the very beginning. Under the canon of Kantian ethics, ethics is not elective but comprehensible rationally.