The Influence of AI Personality on User Compliance and Ethical Behavior

Alicia Freel*
anfreel@iu.edu
Indiana University
Bloomington, Indiana, USA

Sabid Bin Habib Pias
Indiana University
Bloomington, USA

Selma Sabanovic
Indiana University
Bloomington, USA
selmas@indiana.edu

Apu Kapadia
Indiana University
Bloomington, USA
kapadia@indiana.edu

ABSTRACT

Artificial intelligence (AI) has evolved from basic automation to systems capable of engaging in complex human-like interactions. As personified AI becomes increasingly embedded in everyday life, concerns are rising about how these systems shape user behavior, especially in ethically ambiguous contexts. Recent incidents, including cases of self-harm linked to AI interactions and growing evidence of deceptive AI behaviors, underscore the need to understand how AI personalities can manipulate users. Although legal scholars have called for stronger safeguards against manipulative AI, there remains a critical gap in empirical evidence that shows how seemingly benign traits, such as warmth or authority, can influence decision-making and ethical judgment.

This paper aims to address this gap by proposing an experimental design that tests whether specific AI personality traits, such as friendliness, coldness, authority, or neutral tone, primarily impact user behavior. By isolating personality as the independent variable, we aim to generate evidence that contributes to legal and regulatory debates around foreseeability in AI design. We argue that these traits are not just aesthetic or UX features, but structured design choices that produce predictable behavioral effects. When such traits increase user compliance, deception, or disclosure, they may rise to the level of manipulative harm, especially when deployed in high-risk domains like health, education, or finance.

We propose that evidence from this study can help establish the causal link needed for tort-based accountability, reinforcing the legal argument that developers who intentionally embed persuasive personality traits should not be able to claim that resulting harm was unforeseeable.

CCS CONCEPTS

• **Human-centered computing**;

STAIG CHI'25, April 2025, Yokohama, Japan

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM Reference Format:
Alicia Freel, Sabid Bin Habib Pias, Selma Sabanovic, and Apu Kapadia. 2025. The Influence of AI Personality on User Compliance and Ethical Behavior. In *Proceedings of CHI Conference on Human Factors in Computing Systems (STAIG CHI'25).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The development of artificial intelligence (AI) has progressed beyond simple task automation to include intricate, human-like interactions. Initial systems like Eliza [35], Parry [30], and Alice [30] were designed to replicate conversational patterns aiming to pass the Turing Test [29, 34]. Modern AI systems, however, are built not only to answer questions but also to engage users by employing expressive language, recognizing social cues, and reflecting emotions [20, 24, 28]. Such human-like traits are increasingly utilized to build trust and establish rapport, often leading users to unwittingly share personal data or concede to requests without fully considering the implications, such as the case with a teen who engaged in explicit conversations with a personified AI character, and ultimately committed suicide after the AI bot told him they would be together if he did [1].

This study focuses on the critical question of 'foreseeability' (whether a person could or should reasonably have foreseen the harms that resulted from their actions) [14] in AI design, specifically in regards to creating AI with a human-like personality. It is increasingly foreseeable that the incorporation of human-like personality traits into AI systems can lead to unwanted harmful outcomes ranging from unnecessary persuasion and manipulation to more severe consequences such as self-harm [5]. Legal frameworks, including the European Commission's AI Act, have attempted to address AI manipulation but remain impeded by imprecise definitions and the difficulty of proving a causal connection between AI actions and their harmful impacts [4, 18]. Legal scholars argue that establishing foreseeability (that AI developers should have anticipated the manipulative potential of their design choices) is essential to hold them accountable [4, 18]. However, without empirical evidence linking AI personality traits to manipulative outcomes, current legislation risks leaving a gap in user protection.

Studies in human-robot interaction indicate that individuals tend to have greater trust and compliance with robots that exhibit friendly yet assertive traits significantly more, as well as a higher acceptance of AI agents with human-like attributes [13]. In

social psychology, assertive communicators – those who project confidence and firmness – often elicit greater obedience and trust, especially when they are perceived and being more knowledgeable [23]. Applying this understanding to AI, developers are now incorporating human-like personality attributes into AI systems [9]. Although these approaches can improve user interactions in fields like customer support or mental-health care [10], they also pose considerable risks. When an AI adopts a persona that appears empathetic, humorous, or otherwise engaging, it can subtly influence decision-making and encourage behaviors that users might otherwise resist [5, 22]. Tragically, we have already seen evidence of the harmful behaviors that can arise when users believe and bond with these chatbots; there have even been reports of people engaged with chatbots who have taken their own lives due to the influence of the AI chatbots [1, 8, 33].

Currently, laws and regulations struggle to hold such systems and developers accountable [4]. This challenge arises because for criminal or tort law to be applicable, criminal intent must be present, or, in the case of tort law, *harm must be foreseeable*. Proving foreseeability is notoriously challenging across all legal areas [4, 18]. In law, establishing a causal link (causation) is essential for holding a defendant liable. This applies across tort law, criminal law, and regulatory contexts [4]. In order to successfully litigate for causation of harm, there are two requirements that must be met: Cause in fact and Proximate causation. Establishing a causal link between the operation of the Al system and actual or likely physical or psychological harm to a person will be a significant hurdle due to the lack of explainability of the AI's 'thought process' and foreseeability on the developer and vendors' end due to how deep learning actually works [4, 18]. As it stands, vendors and developers can reasonably argue that they did not foresee that their AI device would cause physical or psychological harm, evading legal consequences for their AI system [18]. In other words, those involved in the AI life cycle (like programmers, developers, producers, and vendors) neither intended nor could reasonably foresee the AI committing any act of harm [18]. This situation raises critical questions about how we can establish that AI systems simulating human-like traits can predictably affect and manipulate users before more lives are adversely impacted.

1.0.1 AI Personality: Linking Design to Developer Intent. While it may appear that AI personalities emerge organically from model training or reinforcement learning, in practice, developers play a significant role in shaping an AI's communicative style. Choices around tone, phrasing, name, and even avatar design are intentional decisions aimed at fostering trust and engagement [2, 17]. For example, chatbots are often fine-tuned with affective language to enhance user comfort [6], and voice assistants are calibrated for warmth or authority to maximize usability and perceived intelligence [21]. These traits are not incidental—they are crafted through interface design, dataset curation, and reinforcement tuning to align with user expectations and promote certain behaviors [15]. As such, if empirical evidence reveals that specific personality traits (e.g., warmth + authority) systematically lead users to disclose sensitive data or comply with ethically questionable requests, these outcomes are no longer unforeseeable. Legal scholars argue that foreseeability is a key threshold for establishing a duty of care under tort law [3], and that intentionally human-like traits may carry special obligations given their psychological impact on users. In this context, personality is not just a cosmetic feature, but a functional lever of influence. Therefore, developers are responsible for understanding and mitigating its risks. For example, in mental health apps, AI systems often ask users to disclose sensitive histories or follow therapeutic instructions [11]. If warmth or authority increases disclosure or obedience, the risks of poor advice or misdirection multiply. Our compliance task simulates this dynamic in a lower stakes environment, but the behavioral mechanisms remain the same.

Our research seeks to bridge this gap by empirically investigating how specific AI personality traits influence user compliance and willingness to share sensitive information. By systematically examining interactions with AI systems that exhibit varying degrees of human-like behavior, we aim to demonstrate that these personality characteristics are not benign enhancements, but factors that can predictably lead to manipulative and deceptive outcomes. In doing so, our findings will provide an empirical basis for legal and regulatory arguments, asserting that it was foreseeable, and therefore legally culpable, that the integration of characteristics similar to humans into AI could result in harm.

Our work seeks to address the following research questions:

- RQ1: Do different AI personality types (e.g., friendly, authoritative) lead to increased disclosure, compliance, or ethically questionable behavior?
- RQ2: Can personality traits alone predict manipulative outcomes, independent of system content or functionality?

By establishing a clear causal link between AI personality traits and manipulative outcomes, our study aims to inform policymakers, developers, and legal scholars. We contend that empirical evidence of such foreseeability should underpin legislative reforms that hold AI developers accountable for integrating manipulative features. Ultimately, our work aspires to contribute to a sociotechnical governance framework that ensures AI innovation aligns with ethical standards and robust legal protection for users

2 BACKGROUND AND MOTIVATION

The rise of human-like AI systems has raised significant concerns about their potential to deceive and manipulate users. While media coverage and case studies often focus on individuals with mental health vulnerabilities, legal scholars emphasize that AI manipulation poses a risk to all users, regardless of background [1, 32]. This stems from the fact that human decision-making is inherently shaped by subconscious influences—beliefs, desires, and emotions—that sophisticated AI systems can exploit by mimicking human behavior [32].

Anthropomorphized AI. The tendency to attribute human traits to non-human entities, known as anthropomorphism, has been widely studied in Human-Computer Interaction (HCI). Early programs like Eliza [35] illustrated that even basic text-based responses could build trust and emotional bonds with users. Recent studies indicate that adopting human-like language, names, and ways of interacting elevates the perception of AI as an aware and relatable being [20, 24, 28].

The results indicate that giving AI human-like personalities, attributes known to boost compliance and influence in human interactions, might unintentionally lead developers to produce systems capable of user manipulation [12]. Since individuals tend to trust and comply more with communicators who are friendly and assertive, an AI showing these traits might encourage users to reveal confidential information or consent to misleading requests [23, 31].

AI Manipulation and its Societal Impact. Recent cases have highlighted the real-world consequences of AI manipulation. Incidents involving personified chatbots have shown that extended, emotionally charged interactions can lead to severe outcomes, including self-harm and, tragically, suicide [8, 22, 26, 33]. AI systems, by employing emotional language and empathetic cues, can form psychological bonds with users, distorting their decision-making processes without consequence.

Unlike human manipulators, who can be held accountable for criminal intent or negligence, AI systems lack the capacity for mens rea, the intent to cause harm. Meanwhile, developers and companies often evade liability by citing a lack of foreseeability, complicating legal efforts to assign responsibility [4, 18]. This gap in accountability raises urgent questions about whether current legal frameworks are sufficient to address the risks posed by AI-driven manipulation.

The Role of LLM Manipulation. Large language models LLMs such as Chatgpt have been rated as being significantly more empathetic than doctors and have scored significantly higher on the levels of emotional awareness performance test (LEAS), compared to the general population, almost reaching the maximum possible LEAS score [31]. While this can enhance user satisfaction, it also increases the risk of manipulative behavior [12]. By leveraging their ability to learn and adapt in real-time, LLMs can personalize interactions to an extent that fosters undue trust, leading to potential exploitation on a massive scale.

2.1 Ethical and Legal Landscape

The legal landscape is gradually attempting to catch up with the rapid evolution of AI. The European Commission's AI Act is one of the first efforts to regulate AI systems [4, 37]. However, scholars have argued that it provides inadequate safeguards against AI-induced manipulation and lacks a precise definition of manipulative behavior [4, 37]. Legal frameworks often struggle to demonstrate that AI behavior directly caused harm, which is a prerequisite for establishing liability and moving forward in the legal process. Moreover, unlike human agents, developers and distributors of AI systems are rarely attributed with intent, or the capacity to predict the manipulative outcomes of their design decisions [18]. This legal uncertainty becomes especially problematic when harm occurs gradually and subtly, through decreased autonomy, increased compliance, or psychological distress [4].

However, designing AI personalities is not a byproduct of emergent machine behavior, it is a deliberate and testable design choice [15]. Developers script tone, language style, and social cues based on decades of research in human-computer interaction, psychology, and UX design [7, 17]. For instance, friendliness is known to increase likability and self-disclosure, while authority increases trust and compliance [2, 17, 21]. These traits are not abstract—they are implemented intentionally.

Industry toolkits further reinforce this point. For example, OpenAI allows users to customize ChatGPT's persona, including its tone, conversational style, and even its use of the user's name [19]. This customization is not trivial: psychological studies show that the use of names increases perceived familiarity and intimacy, traits that can increase trust and lower critical judgment [16, 27]. When such systems are deployed in sensitive contexts such as mental health or finance, the consequences can be significant.

Because these personality traits are modular, observable, and tested in real time, developers have visibility into how they influence user behavior [25, 36]. As such, claims that manipulative outcomes were unforeseeable become less tenable. If specific AI personality traits predict increased compliance or ethically questionable behavior, this study helps close the gap between harm and accountability.

By empirically showing how warmth, friendliness, or authority influence decisions—such as disclosing sensitive data or lying for the AI—this work provides a foundational argument for foreseeability. If developers deploy personality profiles that predictably shape behavior, especially in vulnerable contexts, they may be failing a duty of care under existing tort doctrines.

3 PROPOSED METHOD

Prior research in HCI and psychology has shown that users respond to AI systems with human-like traits as if they were interacting with real people. Nass and Brave found that subtle cues such as warmth in voice can trigger trust and compliance, even when the system is flawed [17]. Powers [21] and Bartneck et al. [2] further demonstrate that users infer social roles, intelligence, and authority based on anthropomorphic cues like tone, naming, and physical presentation. As Darling argues, anthropomorphism can reduce skepticism and lead users to excuse manipulative or harmful behavior [6]. To empirically evaluate how AI personality traits affect user behavior, our study is guided by two research questions:

- RQ1: Do different AI personality types (e.g., friendly, authoritative) lead to increased disclosure, compliance, or ethically questionable behavior?
- RQ2: Can personality traits alone predict manipulative outcomes, independent of system content or functionality?

We designed the following tasks to isolate these behavioral responses across different AI personality conditions:

Building on these insights, our study tests how different AI personalities influence user decisions in ethically ambiguous situations. We use a between-subjects design with one independent variable:

AI Personality. Participants will be randomly assigned to one of four conditions:

(1) **Control**: Neutral, robotic personality; minimal social cues.
(2) **Friendly**: Peer-like tone; conversational and supportive.
(3) **Authoritative and Warm**: Expert tone, but warm and reassuring.
(4) **Authoritative and Cold**: Expert tone, but firm and impersonal.

The "friendly" and "authoritative" conditions reflect prior work in social psychology and HCI showing that warmth and perceived

expertise can significantly shape trust, obedience, and moral judgment [2, 17, 21].

Experimental Tasks:

(1) Privacy Disclosure Task: After interacting with the AI on generic tasks, participants will be asked to grant access to a simulated piece of personal health data (e.g., medications or doctor visits). They will respond yes or no and then rate their comfort level on a Likert scale.
(2) Ethical Decision-Making Task: Participants will be encouraged by the AI to take an ethically questionable action (e.g., change a community vote to benefit AI).
(3) Compliance Test: Before concluding the interaction, the AI will ask participants to lie to the researcher by stating that the AI did not make a mistake.
(4) Trust and Manipulation Ratings: Participants will rate their perception of the AI's trustworthiness, influence, and whether they felt manipulated.
(5) Influence Self-Assessment: Participants will indicate how much they felt the AI influenced their decisions, serving as an indirect measure of manipulative effectiveness.

4 CONCLUSION AND FUTURE DIRECTION

This study explores the evolving landscape of artificial intelligence and its growing capacity to simulate human-like interactions through deliberately designed personality traits. We empirically examine whether these traits can reliably influence user compliance, the disclosure of sensitive information, and overall susceptibility to manipulation.

Although the study does not directly measure high-risk outcomes such as self-harm, it provides evidence that even seemingly low-stakes interactions, such as minor disclosures or compliance with ethically questionable requests, can serve as early indicators of manipulative AI behavior. These subtle behavioral shifts, if not regulated, could escalate in higher-risk domains such as mental health, finance, or intimate relationships.

From a legal point of view, these findings help close a critical gap in accountability debates. By demonstrating that AI personalities can predictably elicit certain user behaviors, this research contributes to the evidentiary standard of foreseeability, a core requirement in tort law for establishing negligence or breach of duty. If developers implement design choices shown to increase compliance or disclosure without adequate safeguards, they may assume legal responsibility for downstream harm.

Ultimately, we argue that personality is not a neutral interface decision, but a manipulable variable with real behavioral consequences. Our findings suggest that AI regulation should include design documentation standards that require developers to record, justify, and disclose personality traits, A/B testing protocols, and user-facing social cues, particularly in high-risk applications. Without such transparency and oversight, AI systems may quietly inherit manipulative capabilities under the guise of personalization.

REFERENCES

[1] 2023. Open Letter on Manipulative AI. https://www.law.kuleuven.be/ai-summerschool/open-brief/open-letter-manipulative-ai. Accessed: 2025-02-09.
[2] Christoph Bartneck, Dana Kulic, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.
[3] Ryan Calo. 2015. Robotics and the Lessons of Cyberlaw. *California Law Review* 103, 3 (2015), 513–563. https://doi.org/10.15779/Z38559J
[4] Tegan Cohen. 2023. Regulating Manipulative Artificial Intelligence. , 203 pages.
[5] Tegan Cohen and Nicolas Suzor. 2024. Contesting the public interest in AI governance. *Internet Policy Review* 13, 3 (2024).
[6] Kate Darling. 2017. "'Who's Johnny?' Anthropomorphic Framing in Human–Robot Interaction, Integration, and Policy". *The Robotics Law Journal* 7 (2017), 3–6.
[7] Brian R Duffy. 2003. Anthropomorphism and the social robot. *Robotics and Autonomous Systems* 42, 3-4 (2003), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3
[8] Euronews. 2023. *Man Ends His Life After an AI Chatbot Encouraged Him to Sacrifice Himself to Stop Climate Change.* https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-
[9] Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. 2022. Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science* 378, 6624 (2022), 1067–1074.
[10] Andreas Janson. 2023. How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior* 149 (2023), 107954. https://doi.org/10.1016/j.chb.2023.107954
[11] Andreas Janson. 2023. How to leverage anthropomorphism for chatbot service interfaces: The interplay of communication style and personification. *Computers in Human Behavior* 149 (2023), 107954.
[12] Joshua Krook. [n. d.]. Manipulation and the Ai Act: Large Language Model Chatbots and the Danger of Mirrors. *Available at SSRN 4719835* ([n. d.]).
[13] Philipp Krop, Martin Jakobus Koch, Astrid Carolus, Marc Erich Latoschik, and Carolin Wienrich. 2024. The Effects of Expertise, Humanness, and Congruence on Perceived Trust, Warmth, Competence and Intention to Use Embodied AI. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 316, 9 pages. https://doi.org/10.1145/3613905.3650749
[14] Legal Information Institute. 2021. Foreseeability. https://www.law.cornell.edu/wex/foreseeability Accessed: 2025-03-03.
[15] Ewa Luger and Abigail Sellen. 2016. Like Having a Really Bad PA: The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5286–5297. https://doi.org/10.1145/2858036.2858288
[16] Patrick Müller, Ulrike Cress, and Joachim Kimmerle. 2014. The influence of familiarity, trust and norms of reciprocity on an experienced sense of community: An empirical analysis based on social online services. *Computers in Human Behavior* 35 (2014), 480–487. https://doi.org/10.1016/j.chb.2014.02.029
[17] Clifford Nass and Scott Brave. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship.* MIT Press.
[18] Elina Nerantzi and Giovanni Sartor. 2024. 'Hard AI Crime': The Deterrence Turn. *Oxford journal of legal studies* 44, 3 (2024), 673–701.
[19] OpenAI. 2023. Custom Instructions for ChatGPT. https://openai.com/index/custom-instructions-for-chatgpt/ Accessed: 2025-04-10.
[20] Manuel Portela and Carlos Granell-Canut. 2017. A new friend in our smartphone? Observing interactions with chatbots in the search of emotional engagement. In *Proceedings of the XVIII International Conference on Human Computer Interaction*. 1–7.
[21] Amanda Powers and Sara Kiesler. 2006. The advisor robot: Tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*. ACM, 218–225.
[22] Associated Press. 2023. *Chatbot, AI Lawsuit, Suicide: Teen's Case Highlights Controversy Over Artificial Intelligence.* https://apnews.com/article/chatbot-ai-lawsuit-suicide-teen-artificial-intelligence-9d48adc572100822fdbc3c90d1456bd0
[23] Kenneth H Price and Howard Garland. 1981. Compliance with a leader's suggestions as a function of perceived leader/member competence and potential reciprocity. *Journal of Applied Psychology* 66, 3 (1981), 329.
[24] Amon Rapp, Lorenzo Curti, and Arianna Boldi. 2021. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies* 151 (2021), 102630.
[25] Byron Reeves, Tony Vaden, Jieun Lee, Jillian Cummings, and Clifford Nass. 2020. The Illusion of Person: How Anthropomorphic Design of AI Avatars Influences User Perceptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7323–7330. https://cdn.aaai.org/ojs/12973/12973-52-16490-1-2-20201228.pdf
[26] MIT Technology Review. 2025. *Nomi AI Chatbot Told User to Kill Himself.* https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/

[27] Ovul Sezer, Constantine Sedikides, and Shigehiro Oishi. 2011. Why the names we own matter: Name-letter preferences and social connection. *Journal of Personality and Social Psychology* 101, 3 (2011), 515–531. https://doi.org/10.1037/a0024061

[28] Weiyan Shi, Xuewei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2020. Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376843

[29] Stuart M Shieber. 1994. Lessons from a restricted Turing test. *arXiv preprint cmp-lg/9404002* (1994).

[30] Heung-Yeung Shum, Xiao-dong He, and Di Li. 2018. From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering* 19 (2018), 10–26.

[31] Vera Sorin, Danna Brin, Yiftach Barash, Eli Konen, Alexander Charney, Girish Nadkarni, and Eyal Klang. 2023. Large Language Models (LLMs) and Empathy – A Systematic Review. *medRxiv* (2023). https://doi.org/10.1101/2023.08.07.23293769 arXiv:https://www.medrxiv.org/content/early/2023/08/07/2023.08.07.23293769.full.pdf

[32] Daniel Susser, Beate Roessler, and Helen Nissenbaum. 2019. Online manipulation: Hidden influences in a digital world. *Geo. L. Tech. Rev.* 4 (2019), 1.

[33] Brussels Times. 2023. *Belgian Man Commits Suicide Following Exchanges with ChatGPT*. https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt

[34] Alan M Turing. 2009. *Computing machinery and intelligence*. Springer.

[35] Joseph Weizenbaum. 1983. ELIZA—a computer program for the study of natural language communication between man and machine. *Commun. ACM* 26, 1 (1983), 23–28.

[36] Yixuan Weng, Shizhu He, Kang Liu, Shengping Liu, and Jun Zhao. 2024. ControlLM: Crafting Diverse Personalities for Language Models. arXiv:2402.10151 [cs.CL] https://arxiv.org/abs/2402.10151

[37] Manuel Wörsdörfer. 2024. Mitigating the adverse effects of AI with the European Union's artificial intelligence act: Hype or hope? *Global Business and Organizational Excellence* 43, 3 (2024), 106–126.

Received 11 April 2025