

Signaling Responsibility: The Branding Logic of AI Self-Regulation

As artificial intelligence (AI) becomes integral to corporate identity, technology firms such as OpenAI, Google, Microsoft, Meta, and IBM increasingly adopt voluntary self-regulatory practices—such as publishing AI principles, pausing product releases, establishing oversight boards, and releasing open-source transparency tools. Undertaken without legal obligation, these actions not only address regulatory gaps but also operate as signals of responsibility, leadership, and trustworthiness amid intense public scrutiny. Yet the literature offers limited, systematic evidence on the organizational motives behind such voluntary constraints and on how users interpret these practices.

This thesis aims to address that gap through a comparative case study of selected AI companies that have publicly engaged in ethical self-regulation. Data collection will combine (i) content analysis of corporate communications (press releases, AI principles, transparency reports, blog posts) to trace how firms frame and justify self-regulation, and (ii) semi-structured interviews with customers to elicit perceived signals (e.g., trust, leadership, risk mitigation) and the cues shaping credibility or skepticism.

The goal is to develop a new, empirically grounded typology that captures the dual dynamics of AI self-regulation—as organizational practice and as social signal—thereby advancing understanding of how ethics, reputation, and legitimacy interact in the governance of emerging technologies. The findings should offer actionable insights for scholars and practitioners on how firms mobilize ethics as a competitive and communicative resource.

- *N.B. this thesis topic is only offered in English*