UCLouvain

LOUVAIN
School of Management

Louvain School of Management
and Vytautas Magnus University

CRISIS COMMUNICATION OF BIG TECH FIRMS DURING AI-RELATED ETHICAL SCANDALS

Author's:
Inna Myroniuk

Master's thesis with the view of getting the degrees:
Marketing and International Commerce +
Master 120 credits in Management, Corporate sustainable management

Supervisor's:
Assoc. prof., dr. Miglė Šontaitė Petkevičienė

2025-2026

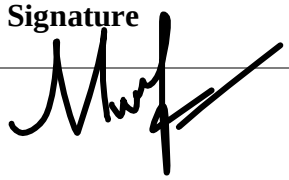
STATEMENTS ON YOUR USE OF GENERATIVE AI

Confirm the following statements:

Statements on responsible use of generative AI	Your answer
The content of my/our master's thesis is my/our original outinput, even if I/we used generative AI to help prepare it.	☑Yes ☐ No
I/we master the theories, tools and methods that I/we use for the thesis.	☑Yes ☐ No
I/we verified that the output presented in the thesis is accurate and valid, and that it contains no hallucination, no false reference, etc.	☑Yes ☐ No
I/we acknowledge that I/we master the final output and that I am/we are able to explain it and answer questions about it.	☑Yes ☐ No
I/we made sure not to provide generative AI tools with access to confidential data or information	☑Yes ☐ No

SIGNATURE

Sign your declaration:

	Author last name, first name(s)	**Date**	**Signature**
1	Myroniuk Inna	28.05.2026	

CONTENT

Inna Myroniuk. DIDELIŲ TECHNOLOGIJŲ ĮMONIŲ KRIZIŲ KOMUNIKACIJA SU DIRBTINIO INTELEKTO ETIKOS SKANDALAIS: Magistro baigiamasis darbas marketingo ir tarptautinės komercijos studijų programoje / Vadovė: doc. dr. Miglė Šontaitė-Petkevičienė / Vytauto Didžiojo universitetas, Ekonomikos ir vadybos fakultetas, Magistrantūros studijos. – Kaunas, 2026. – 91 p.

SANTRAUKA

Tyrimo problema – kaip didelių technologijų įmonės turėtų vykdyti krizių komunikaciją dirbtinio intelekto etikos skandalų metu. **Tyrimo objektas** – didelių technologijų įmonių krizių komunikacija dirbtinio intelekto etikos skandalų metu. **Tyrimo tikslas** – remiantis krizių komunikacijos teorija ir empiriniais tyrimais, parengti įrodymais grįstą krizių komunikacijos sistemą didelių technologijų įmonėms dirbtinio intelekto etikos skandalų metu. **Tyrimo metodai:** mokslinės literatūros analizė, lyginamoji atvejo analizė, kryptingoji kokybinė turinio analizė, internetinė vinetėmis pagrįsta apklausa bei aprašomoji statistinė analizė.

„Google", „OpenAI" ir „Amazon" kokybinė analizė rodo, kad krizių reakcijos skiriasi atsakomybės apibūdinimu, reformų signalais ir etikos plovimo rizika. Kiekybinė apklausa (N = 389) atskleidė, kad respondentai „OpenAI" valdymo krizės atsaką vertino palankiausiai patikimumo, atskaitomybės ir nuoširdumo atžvilgiu, tuo tarpu „Amazon" defleksijos strategija gavo žemiausius įvertinius visais teigiamais kriterijais. Etikos plovimo suvokimas buvo didžiausias „OpenAI" atveju, o tai rodo, kad respondentai vienu metu vertino aiškumą teigiamai ir išliko skeptiški strateginio įvaizdžio valdymo atžvilgiu. Patikimiausi atsakai derina konkrečius taisomuosius veiksmus su stebimais valdymo įsipareigojimais. Išvados: veiksminga krizių komunikacija dirbtinio intelekto etikos skandalų metu priklauso ne tiek nuo abstrakčios etinės retorikos, kiek nuo atskaitomybės, taisomųjų veiksmų, valdymo reformų ir į suinteresuotąsias šalis orientuotos komunikacijos.

Šis darbas yra mokslinio pobūdžio magistro baigiamasis darbas.

Raktažodžiai: atskaitomybė, DI etika, didelės technologijų įmonės, krizių komunikacija, etikos plovimas, suinteresuotųjų šalių pasitikėjimas.

Inna Myroniuk. CRISIS COMMUNICATION OF BIG TECH FIRMS DURING AI-RELATED ETHICAL SCANDALS: Final Master Thesis in Marketing and International Commerce / Supervisor: Assoc.prof., dr. Miglė Šontaitė Petkevičienė / Vytautas Magnus University, Faculty of Economics and Management, Graduate studies. – Kaunas, 2026. – 91.

SUMMARY

Research problem – how to conduct crisis communication for Big Tech firms during AI-related ethical scandals. **Research object** - crisis communication of Big Tech firms during AI-related ethical scandals. **Research aim** - to develop an evidence-based framework for crisis communication of Big Tech firms during AI-related ethical scandals, integrating insights from crisis communication theory and empirical research. **Research methods:** scientific literature analysis, comparative case study analysis, directed qualitative content analysis, an online vignette-based questionnaire, and descriptive statistical analysis.

The qualitative analysis of Google, OpenAI and Amazon shows that crisis responses differ in responsibility framing, reform signals and ethics-washing risk. The quantitative questionnaire (N = 389) indicates that stakeholders rated OpenAI's governance-crisis response most favourably on trustworthiness, accountability and sincerity, while Amazon's deflection strategy received the lowest scores on all positive constructs. Ethics-washing perception was highest for OpenAI, suggesting that respondents simultaneously rewarded clarity and remained sceptical of strategic image-management. The most credible responses combine material corrective action with observable governance commitments. Conclusions: effective crisis communication in AI-related ethical scandals depends less on abstract ethical language and more on answerability, corrective action, governance reform and stakeholder-oriented communication.

This paper is research-oriented final master thesis.

Keywords: accountability, AI ethics, Big Tech, crisis communication, ethics-washing, stakeholder trust.

INTRODUCTION

Relevance of the topic and level of its scientific research. Over the past decade, AI-related ethical scandals have emerged as one of the most consequential governance and reputational challenges that large technology companies face. What makes these situations distinctive is that a crisis – understood in Coombs's (2023) terms as a moment when stakeholders perceive harm, uncertainty, and a violation of their expectations – does not unfold in a vacuum in the AI context. The very features of AI technology compound these pressures considerably: companies must publicly account for systems whose inner workings are technically complex, whose decision-making logic is often opaque, and whose governance choices remain contested – all while operating under conditions of intense public scrutiny and rapid information diffusion across platforms (Cheng et al., 2025; Wen & Holweg, 2024). What follows is a communicative environment that is, in qualitative terms, far more demanding than conventional product-failure or misconduct crises – and one that existing crisis communication frameworks address only partially.

The academic relevance of this topic is grounded in the growing but still insufficiently integrated intersection of three research streams. First, crisis communication scholarship increasingly examines digital and social-mediated crisis environments, where message reception depends not only on the content of what an organisation says but on how that message travels across platforms, is amplified by algorithmic visibility, and is reframed by journalists, activists, regulators, and general publics (Cheng, Wang, & Kong, 2022; Che et al., 2023). The second tradition is AI governance research, which has documented a notable shift in stakeholder expectations: apologies and vague transparency claims no longer suffice. When AI-related harm occurs, affected parties now expect something more demanding – visible answerability, enforceable oversight, and demonstrable limits on organisational power (Novelli et al., 2024; Cheong, 2024). The third tradition, oriented toward questions of legitimacy, points to a risk that has received growing attention in recent years: corporate ethical discourse can actively backfire. When moral language is not supported by structural reform, it tends to be interpreted not as a genuine commitment to accountability, but as a reputational management exercise – what the literature describes as ethics-washing (Hornsey et al., 2024; Van Maanen, 2022; Biddle et al., 2024).

Each of these three traditions is well developed in its own terms. The problem is that they are rarely brought into direct conversation with one another. Research in this space tends to focus on one of three distinct angles: the strategic selection of crisis response messages, the responsible communication of AI capabilities and limitations, or stakeholder trust in digital

governance institutions. What is largely missing is comparative empirical work that connects actual corporate crisis communication texts – the specific language, framing choices, and governance signals that companies deploy – with how stakeholders actually evaluate those communications. This gap is especially significant for Big Tech firms, whose crises are shaped by a combination of platform power, global data infrastructures, and unprecedented public visibility. Communication during a Big Tech AI scandal is not just a reputational issue: it simultaneously concerns governance legitimacy, regulatory risk, and the broader social contract between technology companies and the publics they serve (Huang & Krafft, 2024; Page et al., 2023; Wen & Holweg, 2024). Treating these dimensions separately makes it structurally impossible to account for the full communicative dynamics at play.

The practical relevance of the topic is equally significant. Big Tech companies now routinely operate in high-stakes domains – biometric identification, generative AI, defence-related surveillance, healthcare decision-support – where AI-related controversies extend well beyond brand reputation. Stakeholders do not judge companies on the speed of their communications, but on whether responsibility is acknowledged substantively, corrective actions are independently verifiable, and future harm is structurally constrained. Crisis communication in AI-related ethical scandals has thus become one of the most consequential strategic challenges facing large technology organisations today.

The problem of the research addressed in this paper is how to conduct crisis communication for Big Tech firms during AI-related ethical scandals.

The object of the paper is crisis communication of Big Tech firms during AI-related ethical scandals.

The aim of the paper is to develop an evidence-based framework for crisis communication of Big Tech firms during AI-related ethical scandals, integrating insights from crisis communication theory and empirical research.

To reach the aim the following **objectives** were set:

1. To analyse the theoretical approaches to crisis communication of Big Tech firms during AI-related ethical scandals.

2. To substantiate the empirical research methodology by defining the mixed-method research design and formulating the research parameters.

3. To conduct qualitative comparative case analysis and a quantitative online questionnaire on how Big Tech firms communicate during AI-related ethical scandals and how these responses are evaluated by stakeholders.

4. To formulate evidence-based conclusions and practical recommendations for more credible crisis communication in AI-related ethical scandals.

The paper is **structured** in 3 main parts.

The first part systematises the main concepts of crisis communication, crisis response strategies, digital crisis communication in platform environments, AI-related ethical scandals, accountability and ethics-washing.

The second part, presents the empirical research context, formulates the methodological problem and objectives, reviews relevant methodological approaches, substantiates the mixed-method design, and describes the qualitative and quantitative data collection, processing and reliability procedures.

The third part, presents the qualitative case-study findings, outlines the structure for the quantitative analysis, and integrates the empirical insights into a set of practical recommendations.

Research methods. The study employs a mixed-method design integrating qualitative and quantitative approaches. The qualitative strand applies comparative case study analysis and directed content analysis to official corporate crisis communications produced by Google, OpenAI, and Amazon during three AI-related ethical scandals: Project Maven (2017–2019), the OpenAI governance and copyright controversy (2023–2024), and the Amazon Rekognition facial recognition bias case (2018–2021). Corporate texts were coded against five analytical categories derived from the theoretical framework: responsibility attribution framing, response strategy position on the deflection-to-pre-emption spectrum, the nature of reform and accountability signals, channel and source logic, and ethics-washing risk indicators. The quantitative strand consists of an online vignette-based questionnaire (N = 389) distributed across Europe between 3 April and 3 May 2026, in which respondents were randomly assigned to one of three case conditions and asked to evaluate a standardised crisis-response scenario on four Likert-scale constructs: trustworthiness, accountability, sincerity, and ethics-washing perception. Quantitative data were analysed using descriptive statistics, one-way ANOVA with Tukey and Games-Howell post hoc tests, and Pearson correlation analysis.

Information sources. The literature search was conducted using academic search tools and publisher platforms (e.g., Google Scholar and major journal websites), prioritising peer-reviewed sources in crisis communication, AI ethics/governance, Responsible AI, and Big Tech legitimacy. Author-shared copies were used only when necessary for access, while bibliographic details (journal/publisher and DOI) were verified through the original publication record. Key terms used throughout this study are defined in Annex 1.

The usage of generative artificial intelligence and its tools. Generative AI tools were used in a limited and declared manner for language editing, structure refinement, and translation support. They were not used to generate data, fabricate references, or substitute for

the author's own analysis and interpretation. All cited sources were independently selected and verified by the author (see Annex 2).

1. LITERATURE REVIEW ON CRISIS COMMUNICATION OF BIG TECH FIRMS DURING AI-RELATED ETHICAL SCANDALS

1.1 Definition of crisis communication

Crisis communication is a fundamental element of contemporary emergency management: it supports the delivery of accurate information, helps reduce public distress, and promotes coherent responses aimed at mitigating the impact of critical situations (Cheng, Li, Lee & Liu, 2025). Beyond this functional role, however, the concept has been approached from several different directions in the literature – and understanding the differences between these perspectives matters for how crisis communication is analysed in practice.

The most widely cited view treats crisis communication primarily as an intervention — one designed to limit the negative consequences of a crisis for both the organisation at the centre of events and the stakeholders affected by them (Coombs, 2023). On this reading, communication serves a dual purpose: it addresses the immediate situation directly, but it also works to contain misinformation and prevent organisational trust from eroding further. How effectively a response manages these two objectives can be assessed on what Coombs (2021) describes as a spectrum of accommodativeness, which runs from full acceptance of responsibility at one end to outright denial and deflection at the other.

A second perspective takes a more operational view, defining crisis communication as the collection, processing, and dissemination of information required to address a crisis situation (Coombs, 2010). Where the first perspective focuses on stakeholder impact, this one focuses on organisational capacity – specifically, the ability to gather relevant data from multiple sources under uncertainty and distribute it in ways that support coordinated action and reduce harm (Eisele et al., 2022). In the AI context, this operational dimension becomes particularly demanding. The black-box character of deep learning systems creates real barriers to the kind of meaningful explanation that stakeholders expect, and the structural opacity of many AI failures makes it genuinely difficult to provide a clear account of what went wrong.

A third perspective introduces an interactive dimension, defining crisis communication as an ongoing dialogue between an organisation and its publics before, during, and after a negative event. In digital environments, this has evolved into what can be described as a form of social-mediated dialogue between or within organisations and publics Fearn-Banks, 2002, p. 2). In this context, AI can be treated as a strategic factor supporting crisis communication by

assessing public sentiment, detecting threats, and helping organisations respond to evolving public contestation (Cheng, Wang & Kong, 2022).

The core definitions used in this section are summarised in *Table 1*.

Table 1

Definitions of crisis communication

Definitions	Author
Crisis communication is an intervention designed to lessen the negative effects of a crisis for stakeholders and the organization in crisis.	Coombs (2023).
Crisis communication is defined as the collection, processing, and dissemination of information required to address a crisis situation.	Eisele, Tolochko & Boomgaarden (2022) as cited from Coombs (2023).
Crisis communication is the dialogue between the organization and its public(s) prior to, during, and after the negative occurrence.	Fearn-Banks (2002).

Note: Compiled by the author based on the cited sources.

All three perspectives are connected and work together rather than separately. The intervention lens helps to understand why crisis communication is important: when stakeholders feel that their expectations were violated, even a technical problem can quickly turn into a bigger issue about trust and legitimacy (Coombs, 2023). The operational lens focuses more on what the organisation actually has to do in such situation – collect and share accurate information even when not everything is clear yet (Cheng et al., 2025). This is especially difficult in AI-related cases, because these systems are often hard to explain even for the people who built them, which makes it nearly impossible to give stakeholders the kind of clear answers they are looking for. The dialogic lens adds another layer – it shows that communication is not just about sending a message, because people do not simply accept what a company says. They interpret it, question it, and often form quite strong opinions about the company's character based on how the message sounds (Che et al., 2023).

Taken together, then, the three lenses are not in competition: the intervention frame explains the trigger, the operational frame defines what needs to be made visible, and the dialogic frame describes the contested environment in which any response will land.

Implications for Big Tech AI scandals and the "Double crisis". Such comparison explains why Big Tech AI scandals turn into a "double crisis" which require both reputational

accountability and technology trust. When AI fails and it's not clear why or there isn't enough technical explanation, stakeholders question the company's ability to do its job. People judge the company by how well it can actually understand and fix the problems in its own algorithms — even when those problems are not easy to explain (Wen & Holweg, 2024). But technical failure alone does not fully explain why these situations escalate. The deeper problem appears when the company also fails to meet what Novelli et al. (2024) call answerability — meaning it cannot show that it is being transparent about what went wrong or that its systems are being properly scrutinised.

The reputational side of the crisis works differently. Once a company's response starts to look defensive or out of touch with what society expects, stakeholders stop asking what the company is capable of and start asking what it actually wants to do — and whether its real motivation is responsibility or just protecting itself (Wen & Holweg, 2024). At this point the crisis tends to grow rather than shrink. From an intervention standpoint, this means more and more people feel that their safety and rights are at risk. From a dialogic standpoint, the original controversy gets picked up and reinterpreted across different platforms, and the company's own version of events gets increasingly challenged by journalists, activists, and others.

This is where ethics washing becomes a real risk. If a company's ethical commitments — its principles, statements, advisory boards — are not connected to any real enforcement mechanisms like audits or oversight structures, they start to look like a communication strategy rather than an actual commitment to change. This tension is, in a way, what all three perspectives together point to. For Big Tech firms dealing with AI-related scandals, words are simply not enough — they need to show both that their systems can be governed and that they genuinely intend to govern them.

Why AI-related ethical scandals create a "double crisis" for Big Tech. In AI-related crises, crisis communication becomes "double-layered" because two different reputational assessments are happening at the same time for Big Tech companies. According to Wen and Holweg (2024), stakeholders evaluate companies on two levels at the same time: capability reputation — basically, whether the company can technically identify and fix the problem — and character reputation, which is more about whether their intentions seem honest or mostly self-serving (Wen & Holweg, 2024). When facing pressure on both fronts, companies sometimes try to borrow legitimacy by partnering with trusted external institutions or authorities (Prashantham et al., 2020). Wen and Holweg (2024) also note that firms tend to follow recognisable response patterns — deflection, improvement, validation, and pre-emption — though they often are not well prepared to handle these situations when they actually occur (Wen & Holweg, 2024).

AI governance research adds something important here. Accountability in AI contexts is increasingly understood not just as admitting something went wrong, but as answerability — which requires three specific structural conditions to be in place: authority recognition, interrogation, and limitation of power (Novelli et al., 2024). This matters for crisis communication because stakeholders today expect more than general ethical statements. They want to see actual mechanisms — compliance procedures, reporting obligations, oversight structures, enforcement. Without these, corporate communication risks being read as ethics-washing: using the language of responsibility to manage public perception without actually changing anything (Novelli et al., 2024). In Big Tech AI scandals especially, what makes a response credible is not just what the company says, but whether the governance changes it describes can actually be verified.

To summarise, the literature approaches crisis communication through three overlapping lenses — intervention, operational information management, and interactive dialogue. In the context of AI-related ethical scandals, these three functions cannot really be separated from broader governance expectations. Trust repair depends not only on how well a company communicates, but on whether the accountability signals in that communication are genuinely credible.

1.2 Crisis communication theories and response strategies

To examine the responses of Big Tech companies to AI-related ethical scandals, crisis communication theory should be utilised through various complementary perspectives that elucidate observable corporate reactions.

Crisis response strategies encompass the actions and statements of organisations following criticism, existing on a continuum of receptiveness, ranging from acceptance of responsibility and remediation to outright denial of a crisis (Coombs, 1998). In a defensive way, organisations try to distance themselves from wrongdoing by making excuses, giving reasons, and denying it. This is made easier by the fact that many claims of wrongdoing are vague, which makes it hard for outsiders to judge (Faulkner, 2011). On the accommodating side, organisations take ownership of the problem, offer apologies, and make visible commitments to remedial action (Wen & Holweg, 2024).

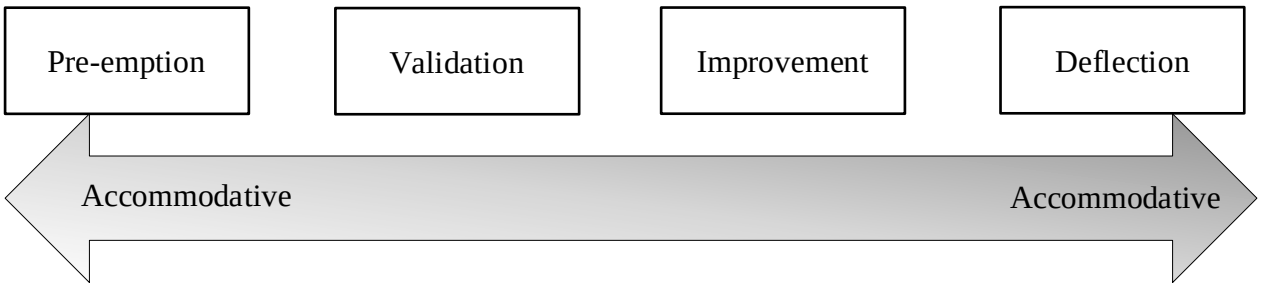
Within the specific context of AI ethical failures, Wen and Holweg (2024) identify four distinct response patterns that map onto this accommodativeness spectrum and reflect the particular characteristics of AI-related controversies

- **Deflection** is the most defensive end of the spectrum, where a firm vigorously defended its interest by shifting responsibility to external actors. In this case, a company strongly defended its interests by shifting blame to outside parties. Companies that use this strategy say that reported wrongdoings are the result of "misuse" by customers or law enforcement. They say that the technology is "not inherently intrusive" but depends on how it is used (Wen & Holweg, 2024).
- **Improvement** involves a partial movement toward the accommodating end of the spectrum. The focus here is on capability reputation — the stakeholder judgement of what the organisation is technically able to do. Rather than disputing the existence of a problem, a company adopting this strategy acknowledges the criticism and responds with concrete technical measures: algorithmic updates to reduce error rates, diversified training datasets to address bias, and similar interventions aimed at the root cause (Wen & Holweg, 2024).
- **Validation** takes a flexible approach to protect character, which is how stakeholders see whether a company's goals and intentions are good or bad. To show that they are trustworthy, companies often try to work with a powerful or influential third party. For example, they might call for government regulation to decide which applications are valid or invite independent auditors to check that their use of technology is ethical (Prashantham et al., 2020).
- **Pre-emption** is the most flexible pattern; it means making moves before criticism reaches its peak. This typology aligns with the double crisis dynamic, wherein firms must convey both technical proficiency (capability reputation) and ethical legitimacy (character reputation) (Wen & Holweg, 2024).

The accommodativeness spectrum discussed in this section is illustrated in *Figure 1*.

Figure 1

Spectrum of company responses to public controversy over facial recognition technology



Source: adapted by the author based on Wen and Holweg (2024).

These response types are analytically useful because they show how firms try to manage two things at once under public pressure: their technical credibility — that is, their demonstrated capacity to identify and correct system failures — and their moral legitimacy, meaning the perception that their intentions and values are appropriate (Wen & Holweg, 2024).

Beyond these response typologies, the literature points to several theoretical frameworks that are directly relevant to understanding how AI-related ethical scandals unfold:

- ***Situational Crisis Communication Theory (SCCT).*** SCCT holds that when a crisis occurs, stakeholders naturally begin attributing responsibility to the organisation involved, and the extent of that attribution shapes both the reputational threat the company faces and the kind of response it should choose (Jong, 2025). In the AI context, this attribution process is rarely straightforward. The opaque, black-box character of algorithmic systems makes it genuinely difficult to determine whether a given failure stems from a technical flaw or a more fundamental design problem — and that ambiguity gives organisations room to shift blame elsewhere rather than accepting accountability (Wen & Holweg, 2024).
- ***Social-Mediated Crisis Communication (SMCC).*** The SMCC model builds on SCCT by incorporating the dynamics specific to social media environments, showing that how a crisis message is received depends heavily on where it comes from and through which channel it travels (Che et al., 2023). A response posted on a company's own platform may land quite differently than the same message circulating through independent users on unaffiliated platforms — with the latter often producing far more sceptical reactions. Related to this, there is evidence that audiences tend to project human-like personality traits onto companies, forming impressions of corporate character based on the tone and framing of official communications (Cheng, Wang, & Kong, 2022).
- ***Legitimacy theory and legitimacy repair.*** Legitimacy theory starts from the premise that organisations do not have an inherent right to operate — they must earn and maintain alignment with the norms, values and expectations of their audiences. From this perspective, corporate texts and public disclosures can be understood as legitimacy-building acts, where language and narrative are used to demonstrate that the organisation is acting appropriately and in accordance with shared social values. When things go wrong, firms may draw on moral positioning and reform narratives to restore both moral and cognitive legitimacy (Hornsey et al., 2024). The risk, however, is that when the discourse is not matched by actual governance capacity or structural change, these responses get read as ethics washing — a form of symbolic communication in which

moral rhetoric serves primarily to manage reputation rather than to drive enforceable reform (Biddle et al., 2024).

The theoretical lenses most relevant to this thesis are compared in *Table 2*.

Table 2

Matrix of theories relevant to AI-related ethical scandals

Theory / approach	**What it explains for the topic**	**Key focus in AI scandals**	**Response implications**	**Main limitation in AI ethical scandals**
SCCT (Situational Crisis Communication Theory)	How stakeholders assign blame and responsibility affects reputational threat, and which response posture is best.	Stakeholders blame AI deployments, data practices, or governance decisions for harm.	Align response strategy with attribution level; prioritise corrective action when accountability is high.	Too "snapshot-like" for ongoing AI crises with repeated revelations and shifting responsibility.
Theory / approach	**What it explains for the topic**	**Key focus in AI scandals**	**Typical response implications**	**Main limitation in AI ethical scandals**
SMCC (Social-Mediated Crisis Communication)	How social media dynamics, like user-generated content, influencers, and platform affordances, affect the meaning of a crisis and the way it affects a person's reputation.	Big Tech crises happen on platforms, and publics, activists, and journalists spread frames. Organizations compete with UGC stories.	Keep an eye on and respond to all channels; expect things to get bigger; talk to people instead of just sending out press releases.	SMCC research shows strong platform/context dependence; generalisations can be difficult without careful case framing and channel-specific analysis.
Legitimacy repair (discursive legitimation)	How organizations "repackage" a crisis through stories and conversations to make it seem legitimate again (for example, by normalizing it, aligning it with moral values, or telling stories about reform that look to the future).	Big Tech might say that AI harms are "unintended," "complex," or "industry-wide," while also talking about commitments, principles, or partnerships with other companies.	Examine persistent legitimization strategies: ethical positioning, external validation, technical determinism, and diffusion of responsibility.	When discourse moves faster than verifiable accountability mechanisms (like audits, enforcement, and transparency), there is a high chance that stakeholders will be skeptical.

Source: compiled by the author based on Wen & Holweg (2024), Cheng et al. (2022), Hornsey et al. (2024)

Viewed through an SCCT lens, all four of these reaction patterns represent different ways of managing how much responsibility gets attributed to the organisation, with each pattern sitting at a different point along the accommodativeness spectrum (Wen & Holweg, 2024). Deflection is the most defensive of the four: rather than acknowledging an internal failure, companies reframe the harm as something caused by external misuse — by customers, law enforcement, or other outside parties — effectively keeping attributed responsibility as low as possible (Wen & Holweg, 2024). Improvement moves closer to acceptance, where the company acknowledges that something technically went wrong and commits to fixing it — through algorithm upgrades, expanded training datasets, or similar interventions — thereby defending what Wen and Holweg call its "capability reputation." Validation takes a different approach, aiming to restore "character reputation" not through technical fixes but through association with credible outside parties, for example by calling for government regulation or bringing in independent auditors to verify that commitments are being honoured (Wen & Holweg, 2024). Pre-emption is the most accommodative pattern of all, involving proactive steps — such as product withdrawals or the early publication of ethical principles — taken before public pressure reaches its peak and before stakeholder attributions of blame have fully solidified (Wen & Holweg, 2024). In AI-related crises specifically, this flexibility matters because the opacity of black-box systems means that causal attributions remain genuinely contested and can shift as new information comes to light (Jong, 2025).

From an SMCC perspective, no crisis response can be evaluated without considering the platform it travels through and the audience that receives it (Che et al., 2023). A message published on a company's own newsroom may be carefully crafted to stabilise the narrative and project control, but once it circulates into third-party spaces, it can easily be reframed as evasive or self-serving regardless of what the original intent was (Opgenhaffen, 2023). Research on the ByteDance crisis illustrates this well: announcements that landed relatively smoothly on affiliated platforms generated far more contested and emotionally charged reactions when the same content reached independent audiences elsewhere (Che et al., 2023). Part of what drives this dynamic is the tendency of audiences to anthropomorphise organisations — reading corporate character from the tone of official statements and labelling companies as bold or cowardly depending on how their communications get interpreted in the wider public arena (Che et al., 2023).

Legitimacy-oriented approaches add another layer by asking whether a given response actually upholds the social contract or just performs the appearance of doing so (Huang & Krafft, 2024). The key finding here is that apologies on their own carry limited weight: stakeholders making trust-repair judgements respond more to concrete reform signals — visible

commitments to behavioural or structural change — than to expressions of regret alone (Hornsey et al., 2024). When corporate rhetoric is not backed by observable commitments such as audits, oversight mechanisms, or reporting obligations, it risks being dismissed as ethics-washing — a rhetorical move that addresses surface concerns while leaving the underlying business logic intact. What stakeholders increasingly demand instead is genuine answerability, which Novelli et al. (2024) define through three structural requirements: authority recognition, interrogation, and limitation of power. Responses that meet these requirements make accountability tangible by connecting discourse to concrete investigation pathways and enforceable constraints that reduce the chance of the same thing happening again (Hornsey et al., 2024).

Taken together, SCCT, SMCC, and legitimacy theory all point to a similar conclusion: AI ethical scandals raise the stakes for responsibility attribution, platform dynamics, and legitimacy repair in ways that standard crisis communication frameworks were simply not built to handle. The four response patterns identified by Wen and Holweg — deflection, improvement, validation, and pre-emption — can each be understood as a strategy for navigating these pressures under conditions of algorithmic ambiguity, where neither the causes of failure nor the right remedies are straightforward.

What makes this especially relevant for empirical analysis is that corporate discourse and actual organisational behaviour can diverge significantly. When they do, ethical language does not repair legitimacy — it actively undermines it. Ethics-washing, in this sense, is not simply a communication failure; it reflects a gap between what a company says it values and what its actual governance structures are capable of enforcing. Ethics councils, published principles, and value statements may create a surface appearance of responsibility, but they function as weak crisis responses when they are not connected to enforceable oversight, independent review, and binding decision-making constraints.

This distinction — between rhetorical reassurance and structural commitment — is therefore central to any empirical evaluation of crisis communication. A response can be considered substantive when it makes accountability visible through mechanisms that can actually be tracked and verified: audits, governance reform, mandatory reporting, external review, and explicit limits on future deployment.

The legitimacy risk connected to purely symbolic responses has attracted sustained scholarly attention. Van Maanen (2022) argues that a significant part of the tech sector's engagement with AI ethics amounts to little more than a way of deflecting external pressure without actually changing core practices. The principles-based model of ethics, Van Maanen argues, has shown a notably limited capacity to redirect or constrain the dominant direction of

Big Tech (Van Maanen, 2022). What fills the gap instead are performative structures — ethics offices and advisory bodies that function primarily as reputational management tools rather than as genuine sites of governance authority. Google's Advanced Technology External Advisory Council (ATEAC) is frequently cited as a concrete example: composed of volunteers whose recommendations were non-binding and could be freely disregarded by R&D teams, it lacked both the centrality and the power needed to influence actual business decisions, and was disbanded within days of its announcement.

The empirical implication of all this is that structural commitment and ethics rhetoric need to be treated as analytically separate categories. Meaningful accountability — what Novelli et al. (2024) call a relation of answerability — requires more than just published principles. It depends on fulfilling three structural conditions: authority recognition, interrogation, and limitation of power, put into practice through the goals of compliance, reporting, oversight, and enforcement.

In summary, combining SCCT and AI governance frameworks suggests that companies have a range of strategic options when facing AI-related ethical scandals, from the most defensive postures to the most proactive ones. But the effectiveness of any of those options ultimately depends on whether the reform signals they contain are seen as genuinely credible or as cover for business-as-usual. The following section moves to the specific digital environment in which Big Tech crises unfold — one where the construction, interpretation, and contestation of crisis responses happens across owned channels, platform ecosystems, and accountability artefacts such as newsrooms, transparency hubs, and compliance reports.

1.3 Digital crisis communication of Big Tech firms

Digital crisis communication in Big Tech operates quite differently from crisis communication in conventional organisational settings. What sets these companies apart is not just their size or public visibility — it is the fact that they often own or govern the very platforms on which controversies about them are debated, amplified, and interpreted (Huang & Krafft, 2024). This creates a quite unusual situation — crisis responses here go far beyond press statements. Big Tech companies also use what could be called platform-facing accountability practices: newsroom posts, transparency reports, compliance submissions, and governance tools for users, all of which are part of how these companies communicate during a crisis.

One important aspect of this is the reliance on transparency-based self-regulation, which has its roots in wider internet governance discussions (Che et al., 2023). Whether this actually works depends on whether the transparency produced is genuinely useful — meaning,

whether the released information allows different actors to hold the platform accountable in a real way (Mehta & Erickson, 2022). When the information is too limited or too carefully selected, it tends to have the opposite effect, leading to accusations that users and citizens have no real way to understand the rules these platforms operate under (Mehta & Erickson, 2022).

Another dimension is how algorithmic visibility and fast-moving online attention shape the crisis communication environment. Some content gets amplified by recommendation algorithms — pushed into timelines and trending sections in ways that speed up public attention (Opgenhaffen, 2023) — so crisis responses in Big Tech cannot really be separated from how the platform itself is designed. What gets seen, what spreads, and how people understand an incident is partly about platform design, not only about what the company decides to say.

Third, the responses companies produce in these environments can be seen as a kind of performative platform governance, or what Huang and Krafft (2024) call stage management. Looking at Meta's practices, they show that companies keep a clear separation between front-stage data relations — the visible side of how data is handled, like interface changes or actions against policy violations — and back-stage data relations, which are the recommendation and advertising systems that bring in most of the revenue but stay mostly hidden from public view (Huang & Krafft, 2024). For crisis communication, this matters because it means that platform design decisions and governance artefacts are themselves part of how legitimacy is managed, not just background contex.

Fourth, Big Tech firms have built accountability infrastructures — transparency reports, advertising libraries, compliance dashboards — that have become standard in crisis communication. However, research keeps raising doubts about what these tools actually achieve. A review of platform reporting under the EU Code of Practice on Disinformation found that compliance was incomplete and that qualitative reporting often lacked substance, while quantitative data was frequently inconsistent (Mündges & Park, 2024). Research on the Facebook Ad Library found similarly that despite the amount of information available, the level of detail on spending, reach, and targeting is not enough to allow real external oversight (Mehta & Erickson, 2022). This suggests these tools should not be seen as neutral proof of accountability, but rather as strategic outputs meant to manage public reaction and reduce the chance of regulatory action (Huang & Krafft, 2024).

Finally, Big Tech crises are cross-platform and socially mediated by nature, which means official statements rarely stay in the channels where they were first published (Cheng et al., 2025). The ByteDance case shows this clearly — announcements that appeared on affiliated platforms like Toutiao were significantly reframed, and received much more negatively, when they spread to independent spaces like Sina Weibo. Different audiences, seeing the same

content through different channels, formed very different impressions of the company, often building an image of its character — whether tough, cowardly, evasive, or principled — based on the tone of what they read (Che et al., 2023). The practical conclusion from the literature is that companies cannot afford to be passive in the most debated public spaces; a proactive, multi-platform approach is needed to avoid losing control of the narrative entirely.

What connects all these dimensions is a recurring gap between what accountability artefacts claim to be and what they actually do. Front-stage governance activity creates the appearance of transparency and control, while back-stage systems — the algorithms and data flows that are central to these companies' business models — remain largely hidden and unaccountable (Huang & Krafft, 2024). Altogether, digital crisis communication in Big Tech needs to be understood as a layered system made up of official messaging, cross-platform circulation and contestation, and accountability artefacts — each of which can either strengthen the credibility of a response or increase public scepticism when it looks like disclosure for its own sake rather than real accountability.

1.4 AI-related ethical scandals as a type of crisis

AI-related ethical scandals are a specific and distinct type of crisis. They are built around public controversies about AI systems that touch on social norms, rights, and expectations that people broadly share — and they rarely happen as one clear, contained event, but rather develop as ongoing conflicts that spread across multiple areas at the same time (Venturini, 2010). Looking at the research literature, there is a recognisable pattern in how these controversies unfold: a technical problem only becomes a real crisis when the behaviour of a deployed AI application is seen as violating social norms and values, and when this starts generating sustained public criticism. There are two elements that drive this shift. The first one is normative: when an AI system is seen as having broken the social contract — meaning it has crossed the explicit and implicit expectations that stakeholders hold about how organisations are supposed to conduct themselves — the problem stops being just a technical defect and gets reframed as a legitimacy conflict (Wen & Holweg, 2024). The second element is relational: once criticism moves outside of internal company processes and reaches the wider public sphere, the organisation faces what Floridi et al. (2022) describe as an ethical failure of AI — a situation where the technology violates social norms and generates public censure (Floridi et al., 2022). It is also worth noting that these failures are partly constructed after the fact: ethical principles for AI are frequently developed as a reaction to high-profile controversies, such as Google's involvement in military drone targeting or Amazon's facial recognition bias, rather

than being established in advance of them. For a failure to actually constitute a genuine crisis, it needs to reach a certain level of visibility — generating criticism across media outlets, social platforms, and policy discussions (Wen & Holweg, 2024). What this means for crisis communication is that Big Tech firms are never simply reacting to a fixed, singular event; they are responding to a dynamic and contested rhetorical situation, where journalists, regulators, and expert commentators are all simultaneously evaluating and judging the corporate narrative, and any misalignment with how the public is interpreting the scandal can quickly produce new cycles of criticism.

1.4.1 Core scandal themes and stakeholder expectations

Although AI scandals vary, they often cluster around recurring ethical concern-types that map directly onto stakeholder expectations. In their analysis of facial recognition technology authors observe that failures commonly relate to aspects of privacy, bias and safety potential malicious use of the technology (Wen & Holweg, 2024). Across the broader Big Tech landscape, these concerns map onto four primary controversy pillars:

1. **Bias and discrimination:** refers to unintentional inequities produced by data or model design choices, which lead to unequal outcomes in consequential settings such as hiring or healthcare. When a scandal centres on this pillar, the organisation's communication is evaluated primarily on whether it can demonstrate that the system has actually been corrected — not just reframed or re-described (Wen & Holweg, 2024). They are looking for concrete technical steps, like algorithm updates or more diverse training data, that directly address the problem. Saying "we take this seriously" is not enough — companies need to show actual evidence of testing and monitoring (Wen & Holweg, 2024).
2. **Privacy and surveillance:** concerns arise when AI systems intrude on individual autonomy through unauthorised data collection or large-scale monitoring of people. In these types of controversies, the communicative challenge shifts away from capability and toward character: stakeholders become less concerned with how the technology performs technically and more focused on whether the company's intentions can actually be trusted (Wen & Holweg, 2024). What matters most is specific, credible information about data flows — what exactly was collected, who had access to it, and what protections are now in place (Wen & Holweg, 2024). Legitimacy in this context depends on making these practices genuinely understandable to people, reducing ambiguity rather than replacing clear data governance commitments with vague ethical

declarations (Cheong, 2024). Effectiveness depends on whether the organisation can provide a real right to explanation — one that actually accounts for the logic behind automated decisions, rather than retreating into general statements about responsibility (Cheong, 2024).

3. **Safety and societal harm:** covers the risks that autonomous systems can create, whether through direct physical harm or more indirect harms like the distortion of public discourse through deepfakes. When a scandal is framed around safety, crisis communication has to address both the failure of the system itself and the broader risk that the technology will be maliciously exploited (Wen & Holweg, 2024). Deflecting responsibility onto end-users or law enforcement who misuse the technology is a common response, but it tends to be insufficient: stakeholders expect the organisation to explain what preventive controls exist and what limits have been placed on how the technology can be used (Wen & Holweg, 2024). The most credible responses in this category are pre-emptive — companies that voluntarily withdraw or pause controversial products before any regulatory requirement forces them to do so send a qualitatively stronger signal about their genuine commitment to safety and human rights (Wen & Holweg, 2024).
4. **Transparency and accountability gaps:** emerge from the black-box character of many AI systems, which makes it structurally difficult to trace the causes of failures or assign responsibility to specific actors — what the literature describes as the many hands problem. When a scandal is driven by this opacity, stakeholders are not primarily looking for an apology; they are looking for answerability (Novelli et al., 2024). Making accountability visible under these conditions requires meeting three structural conditions: authority recognition, interrogation, and limitation of power (Novelli et al., 2024). Crisis communication in this context becomes a test of governance reform — stakeholders are scanning for evidence of empowered oversight, audits, reporting mechanisms and enforcement capacity, rather than abstract ethical language deployed to deflect without changing how the organisation actually operates.

Across all four categories, the pattern is the same: saying sorry is not enough. People want to see that the company can actually be governed and held responsible. Cheong (2024) puts it clearly: transparency helps people understand how AI systems affect their lives, while accountability gives them a way to assign responsibility and get redress (Cheong, 2024).

This means that crisis communication in AI scandals can no longer be evaluated purely on reputational terms. It is increasingly assessed against a governance standard — the capacity to provide meaningful explanation pathways and credible, enforceable reform commitments

(Novelli et al., 2024). When companies fall short of this standard, their responses tend to invite the charge of ethics-washing: that published guidelines and ethical commitments function primarily as reputation protection rather than as a genuine constraint on business conduct (Wen & Holweg, 2024).

1.4.2 Structural complexity and the challenge of responsibility attribution

One of the defining features of AI-related ethical scandals is how structurally complex it is to figure out who is actually responsible for unintended outcomes. AI systems are frequently built on deep neural networks that work as black boxes — their inner workings are often simply impossible for humans to meaningfully inspect. Novelli, Taddeo, and Floridi (2024) point out that while accountability is widely recognised as a cornerstone of AI governance, it tends to be defined in a vague and imprecise way in practice — and this is a problem that comes directly from how sociotechnically complex these systems are. To make the concept more useful, they offer a more rigorous definition: accountability understood as a relation of answerability, which requires three conditions to be satisfied — authority recognition, interrogation, and limitation of power — and which is assessed against four practical goals: compliance, reporting, oversight, and enforcement.

These distinctions matter for crisis communication because they help clarify what stakeholders are actually looking for when a Big Tech AI failure happens. They are not simply asking for a technical explanation of what went wrong; they want to know who has the standing to hold the firm accountable (authority recognition), how the harm will be properly investigated rather than quietly managed away (interrogation), and what structural constraints will be put in place to make the same thing less likely to happen again (limitation of power). The difficulty is that the inherent opacity of many AI systems makes all of these forms of accountability genuinely hard to demonstrate — outcomes are difficult to predict, causes are difficult to isolate, and the chain of technical decisions that led to a failure is rarely visible or transparent (Tsamados et al., 2022). This opacity creates something that resembles what is known as the many hands problem: when so many actors, systems, and technical layers are involved, it becomes practically impossible to assign individual responsibility with any real confidence (Cooper et al., 2022). Companies are able to take advantage of this ambiguity — attributing failures to misuse by external parties rather than to their own internal design choices — which creates a persistent and often unresolved tension between corporate deflection and stakeholder demands for identifiable, enforceable accountability.

1.4.3 Ethics-washing as a backlash pathway

AI ethical scandals keep exposing the same specific and recurring legitimacy problem: the gap between what companies publicly commit to and how they actually behave. One pattern that has attracted a lot of both scholarly and public attention is the tendency of technology companies to publish ethics guidelines not as a genuine expression of their organisational values, but as a mechanism for absorbing criticism and avoiding external scrutiny. Principle-centred approaches of this kind have shown very limited ability to actually redirect the practices of large technology companies in any meaningful way. What many observers now describe as Big Tech's embrace of AI ethics looks, on closer inspection, much more like a public relations strategy — a way of appearing responsible without making the structural changes that genuine responsibility would actually require. In crisis communication terms, this kind of signalling can backfire badly: when audiences sense that they are being managed, value statements tend to deepen distrust rather than repair it.

Research into how ethics actually functions inside organisations reinforces this picture. Ethics offices are frequently positioned in ways that limit their real authority — they sit outside the decision-making processes that govern research and product development, which means their recommendations can be, and often are, simply set aside. For this reason, corporate responses to AI scandals cannot be evaluated on the basis of rhetoric alone; the meaningful question is whether they involve real structural commitments, including independent audits, formal oversight mechanisms, transparent reporting, and enforceable accountability standards.

It is also worth stepping back and acknowledging the broader dynamic at work here. Crisis communication in the Big Tech AI context is not a discrete event that can be handled with a well-crafted statement and then considered resolved. It is an ongoing process shaped by layered and often competing pressures. Companies in these situations have to navigate two distinct reputational challenges at the same time: one technical — whether they are seen as capable of diagnosing and actually fixing the problem — and one moral — whether they are seen as genuinely motivated to act responsibly rather than just trying to protect their market position.

No single theoretical framework is enough to account for the full complexity this creates. Some approaches are useful for understanding how audiences allocate blame and what that means for reputational risk, but they tend to treat crises as bounded events rather than the prolonged, multi-front controversies that AI scandals typically turn into. Others show how the same message gets received differently depending on the platform and medium, and how corporate personality gets projected onto the tone and style of communication. Research on

trust repair points toward a conclusion that is by now fairly well established: recovering from an ethics-related failure requires more than an apology — it requires credible, independently verifiable evidence that something has genuinely changed.

The Big Tech environment also introduces further complications that are specific to this context. These companies do not just face scrutiny from outside — they also own or control many of the platforms on which controversies develop, and they produce the accountability instruments —transparency reports, compliance dashboards, governance documentation — that are supposed to demonstrate responsible conduct. In practice, these instruments frequently function more as tools for 125 managing public perception than as genuine mechanisms of accountability; they provide information that is often too aggregated, too vague, or too self-reported to allow for independent verification.

AI-related ethical scandals, looked at through the lens developed in this chapter, represent a distinct crisis category with three defining features. First, they are norm-based public controversies: they arise when AI systems behave in ways that violate broadly shared social expectations and attract sustained, cross-platform criticism. Second, they are characterised by socio-technical opacity: the internal workings of AI systems are difficult to interpret and explain, which makes responsibility hard to attribute and allows accountability to become so dispersed across layers that it effectively disappears. Third, they trigger elevated governance expectations, with the public and regulators expecting formal, enforceable accountability mechanisms rather than voluntary commitments. When companies fail to meet these expectations, their communications risk being dismissed as performative — empty gestures that substitute reputational signalling for genuine oversight in an environment where AI ethics commitments remain largely unenforceable.

2. EMPIRICAL RESEARCH METHODOLOGY OF CRISIS COMMUNICATION OF BIG FIRMS DURING AI-RELATED ETHICAL SCANDALS

This section's main goal is to outline the empirical research's methodological framework in order to guarantee the study's legitimacy, consistency, and analytical validity. In order to accomplish this goal, the section presents an overview of the research object and research situation, examines pertinent prior empirical studies, develops and supports the research questions and hypotheses, and defends the chosen research methodologies. A comparative mixed-methods approach forms the basis of this thesis' empirical research design. It combines a quantitative vignette-based survey to look at stakeholder assessments of various crisis response components with a qualitative multiple-case analysis of Big Tech companies' official crisis communication during AI-related ethical scandals. Building on the literature review, the section formulates two research questions and three hypotheses, identifies the main analytical criteria, and explains the logic of integrating qualitative and quantitative strands.

2.1 Research methodology

The empirical research focuses on three Big Tech firms – Google, OpenAI and Amazon – and on three publicly documented AI-related ethical scandals that differ in controversy type but are comparable in their legitimacy implications. The selected cases cover military use and internal protest (Google Project Maven), governance breakdown and copyright controversy (OpenAI), and algorithmic bias in facial recognition deployed for law-enforcement purposes (Amazon Rekognition). These cases were selected because they combine public criticism, reputational pressure, governance concerns and a visible corporate communication response.

Overview of the research object: selected organisations

The three companies selected for this study — Google, OpenAI, and Amazon — represent distinct but comparable positions within the Big Tech landscape, each combining significant AI development capacity with substantial public visibility and governance responsibilities.

Google (Alphabet Inc.) is one of the world's largest technology companies, operating across search, cloud computing, advertising, and AI research. Through its DeepMind and Google Brain divisions, Google has been at the forefront of AI development for over a decade

and was among the first Big Tech firms to publish a formal set of AI governance principles. Its scale, global reach, and dual role as both an AI developer and a platform operator make it a central actor in debates about responsible AI deployment.

OpenAI is an AI research and deployment company founded in 2015, originally structured as a non-profit with the stated mission of developing artificial general intelligence for the benefit of humanity. Following a transition to a capped-profit model and a major strategic partnership with Microsoft, OpenAI became one of the most prominent and commercially influential AI companies in the world, primarily through the public release of ChatGPT and the GPT model series. Its unusual governance structure — combining a non-profit board with a commercial subsidiary — makes it a particularly relevant case for examining tensions between mission-driven discourse and commercial behaviour.

Amazon (Amazon Web Services, AWS) is the world's leading cloud computing provider and one of the largest technology conglomerates globally. Through AWS, Amazon has developed and commercially deployed a range of AI-powered services, including its Rekognition facial recognition system, which has been sold to law enforcement agencies across the United States. Amazon's position as both an AI developer and a major infrastructure provider for other organisations gives its governance and communication decisions outsized consequences for the broader AI ecosystem.

Together, these three organisations offer a comparative basis for examining how Big Tech firms with different governance structures, business models, and public identities navigate AI-related ethical controversies.

The situation under investigation is characterised by two interrelated features. First, AI-related ethical scandals are not limited to technical malfunction; they trigger public debate about accountability, fairness, transparency, human rights and corporate power. Second, the communication environment is strongly digital and platform-mediated, meaning that official statements are continuously compared with journalism, activist framing, expert commentary and user reactions. This makes the research object suitable for a mixed-method design that combines qualitative analysis of actual crisis texts with quantitative evaluation of stakeholder perceptions.

The need for this empirical research is justified by the lack of studies that connect these two levels in one analytical design. Previous research has analysed either the content of corporate communication or its effects on audiences, but less frequently both together in the context of AI-related ethical scandals. This study addresses that gap directly, by looking at how Big Tech firms frame responsibility, corrective action and governance reform, and by examining how these elements shape stakeholder perceptions of trustworthiness, legitimacy,

accountability, sincerity, corrective-action credibility and ethics-washing.

The mixed-method design is therefore not merely a methodological convenience — it is substantively required by the research problem itself. The qualitative strand captures how organisations actually construct their crisis communications, while the quantitative strand examines how stakeholders receive and evaluate those same types of messages. Bringing both strands together makes it possible to study crisis communication as something that is simultaneously produced by organisations and interpreted by audiences.

Research problem is how to conduct official crisis communication of Google, OpenAI and Amazon during selected AI-related ethical scandals.

The research object is the official crisis communication of Google, OpenAI and Amazon in selected AI-related ethical scandals.

The aim of the research is to identify recurring response patterns in the selected cases and to evaluate how stakeholders perceive the credibility and substance of key crisis-communication elements.

To reach the aim the following **objectives** were set:

1. To define the empirical research context, select comparable cases and establish the qualitative analytical categories.
2. To formulate qualitative research assumptions, quantitative hypotheses and evaluation criteria grounded in the theoretical analysis.
3. To justify and apply a mixed-method design that combines comparative qualitative case analysis with an online vignette-based questionnaire.
4. To compare the qualitative and quantitative findings and derive empirically grounded conclusions and practical recommendations.

To achieve these objectives, the study draws on two complementary research methods: qualitative and quantitative. The qualitative component involves a case study analysis of three AI-related ethical scandals, selected to capture variation in controversy type and corporate response. The quantitative component consists of an online questionnaire targeting active users of AI-powered technologies.

The research logic of this study is based on integrating two methods to look at crisis communication of Big Tech firms during AI-related ethical scandals from both sides of the communication process. Used together, they allow for a better understanding of how corporate crisis responses are strategically constructed and how they are ultimately judged by the people who receive them. Qualitative research is needed here because understanding how tech firms frame their responses to AI ethical controversies requires close reading of the actual texts they produce. The qualitative strand is built on a comparative analysis of official corporate crisis

communication materials — including executive statements, newsroom posts, press releases, and public policy updates issued in response to AI-related scandals. This approach makes it possible to examine how firms handle responsibility attribution, acknowledge wrongdoing, communicate corrective action, signal governance reform, and position themselves relative to stakeholder expectations. In situations where legitimacy and public trust are under pressure, this kind of textual analysis reveals a lot about the substance, tone, and strategic intent of crisis communication.

The quantitative strand is designed to complement this by measuring how stakeholders actually perceive specific crisis-response elements. Participants in an online vignette-based survey were asked to evaluate brief, standardised crisis-response scenarios that varied in terms of transparency framing, corrective action, accountability signals, and stakeholder engagement. This approach allows for a more systematic assessment of how different communication choices influence perceptions of trustworthiness, sincerity, legitimacy, corrective-action credibility, and ethics-washing, and makes it possible to identify which message features tend to produce more or less favourable evaluations.

Combining these two methods follows a triangulation logic that strengthens both the validity and the analytical depth of the findings. The qualitative data shows how Big Tech firms communicate during AI-related controversies; the quantitative data helps explain how those communications are actually received. Together, they make it possible to connect organisational communication practices with audience-level judgements within a single empirical design.

Using both approaches also allows for a more thorough examination of the research object overall. Qualitative analysis is better suited to capturing complexity, framing variation, and strategic differences across firms, while quantitative analysis offers a more structured basis for comparing stakeholder responses to specific message elements. The mixed-method design therefore supports a more complete and analytically balanced investigation of crisis communication in AI-related ethical scandals.

Aim of the research and justification for its need and relevance: This study examines the crisis communication strategies used by major technology companies during AI-related ethical scandals and assesses how stakeholders evaluate those strategies. There are several reasons why this topic deserves empirical investigation.

First, AI-related ethical controversies — covering issues such as bias, discrimination, privacy violations, surveillance, lack of transparency, and broader societal harm — have become a regular part of public debate. When these situations arise, Big Tech companies are expected not only to manage their reputations but to demonstrate genuine responsibility and a

credible commitment to change. Understanding how these companies communicate during ethically charged crises is therefore important both for research and for practice.

Second, while researchers have paid increasing attention to crisis communication, AI governance, and digital ethics, these topics are often studied separately. Earlier research has looked at crisis communication via social media, ethical problems in AI, responsible AI discourse, and stakeholder trust. What is still missing is empirical work that brings these threads together by linking actual corporate communications to the perceptions they generate among audiences. This study addresses that gap through a combination of primary survey data and systematic analysis of real corporate crisis texts.

Third, stakeholders are paying increasing attention not just to what technology companies say, but to whether their communications reflect genuine accountability or serve mainly to protect the company's image. In AI-related controversies, this distinction matters a great deal — corporate communication can be read either as a serious attempt to acknowledge harm and commit to change, or as a reputational exercise that avoids meaningful reform. That distinction is particularly consequential for Big Tech firms, given the scale of their social influence and governance responsibilities.

The reason for combining qualitative and quantitative methods is that they complement each other well. Qualitative analysis makes it possible to examine how firms construct their crisis narratives and manage responsibility attribution. Quantitative analysis provides structured, comparable data on how different message characteristics are evaluated by audiences. Together, these methods allow for a fuller picture of the research subject and make it possible to draw empirically grounded conclusions about responsible communication during AI-related crises.

2.2 An overview of research methods used by other authors

To understand what methods are available for studying this kind of topic, it helps to look at how previous researchers have approached related questions. While no existing study covers exactly the same ground as this thesis – combining case analysis of Big Tech crisis communication with stakeholder perception data – several areas of prior research are methodologically relevant.

The first is empirical research in the social media crisis communication (SMCC) tradition. Work in this area has relied heavily on quantitative methods — particularly content analysis and experimental designs — to examine both how crisis messages are structured and how they affect audiences (Cheng et al., 2022). For much of this research, Situational Crisis

Communication Theory has served as the primary theoretical anchor, with reputational outcomes as the key dependent variable. More recently, however, scholars have started arguing that this focus is too narrow, and that crisis communication research should be asking broader questions about societal-level effects rather than just corporate image (Page et al., 2023). This shift is methodologically relevant here because it justifies combining content analysis of corporate texts with survey-based measurement of how stakeholders actually interpret those texts.

A second relevant area concerns the dynamics of online crises and digital backlash. Research here has shown that how a controversy develops online is partly shaped by platform conditions — the rapid, public nature of Twitter, for instance, creates pressures that simply do not exist in the context of a formal press release (Qu et al., 2023). Studies have also looked at specific message features, including whether a conversational tone makes corporate responses seem more or less credible. One relevant finding is that a conversational human voice appears to be more effective when the message is specific and concrete, but not when it is vague or generic — a distinction that is directly relevant to this thesis, given that the crisis responses under analysis vary considerably in their degree of specificity (Huang & Ki, 2023).

Third, empirical AI ethics research has moved toward examining what Wen and Holweg (2024) call "AI ethical failures." Their comparative study of facial recognition controversies is particularly relevant here: it identified four organisational response types ranging from Deflection at the defensive end to Pre-emption at the accommodative end, and this typology forms the primary analytical framework for the qualitative case analysis in the present thesis (Ray et al., 2025).

A fourth relevant strand looks at how major technology companies communicate about responsible AI (RAI) and corporate data responsibility (CDR) in public-facing documents (Lim et al., 2025). Comparative textual analysis of these communications has identified what some scholars call a "moral gap" — a systematic discrepancy between the ethical commitments companies actually articulate and the expectations of international bodies or civil society organisations (Jang & Plaisance, 2025). One finding with direct implications for questionnaire design is that stakeholders tend to weight concrete commitments in areas like privacy, safety, and security more heavily than abstract value statements when forming judgements about a company's overall ethics (Lim et al., 2025).

Fifth, a number of studies have looked at the relationship between transparency and legitimacy specifically in the context of AI communication, and the findings are somewhat counterintuitive. Schilke and Reimann (2025) describe a transparency dilemma in which revealing AI usage can actually reduce perceived legitimacy, because it leads audiences to

question whether their autonomy has been constrained (Schilke & Reimann, 2025). What appears to matter more than disclosure itself is whether the company can demonstrate genuine accountability — a finding that directly informs the decision in this thesis to treat transparency claims and accountability signals as separate constructs in the questionnaire (Yoganathan et al., 2025).

Finally, critical research on ethics-washing and symbolic ethics communication provides both theoretical grounding and methodological precedent for one of the central analytical categories in this study. Rather than assuming that corporate ethical commitments are automatically credible, this research looks at the conditions under which stakeholders find them genuine or merely performative. This is directly relevant because one of the main things this thesis wants to understand is whether audiences can tell the difference between substantive and symbolic crisis responses.

What is missing from this existing literature, and what this thesis aims to contribute, is an integrated empirical design that connects actual corporate crisis communication texts with stakeholder-level assessments of those texts. Most previous work has done one or the other – either analysing corporate communication qualitatively, or measuring stakeholder perceptions quantitatively – but not both in the same study, in the context of AI-related ethical scandals. This gap provides the main justification for the mixed-method approach used in this thesis..

2.3 Formulation and substantiation of research statements

Research assumptions and hypotheses serve a structuring function in empirical work: they translate the theoretical framework into testable propositions and set the criteria against which findings will be evaluated. The statements below were developed on the basis of a systematic review of literature on crisis management and AI-related ethical controversies in Big Tech, and they guide both the qualitative and quantitative strands of the study.

Research assumptions for qualitative research:

RA1: Big Tech firms' official crisis communication during AI-related ethical scandals reveals recurring and distinguishable response patterns rather than random or purely ad hoc reactions.

This assumption rests on the observation that organisations facing comparable kinds of reputational pressure tend to draw on similar communicative repertoires. Rather than responding in an entirely improvised or unpredictable way, firms typically gravitate toward recognisable strategies — whether that involves denying responsibility, announcing technical fixes, seeking external validation, or setting proactive ethical boundaries before public pressure

peaks. Wen and Holweg (2024) showed this quite directly in their study of facial recognition controversies, where they identified four distinct and recurring response types across different companies. The present study assumes that a similar kind of pattern will be visible in the three cases analysed here, even though the scandals themselves differ in type and context.

RA2: In the crisis communication of Big Tech firms, responses that emphasise concrete corrective action are more likely to appear substantive than responses dominated by abstract ethical rhetoric.

This second research assumption is grounded in the distinction between substantive and symbolic crisis communication in AI-related ethical scandals. Recent scholarship suggests that stakeholders increasingly scrutinise whether corporate communication reflects actual responsibility and reform, or whether it mainly serves reputational self-protection through ethical rhetoric. In the qualitative part of this study, this assumption will be operationalised through directed content analysis of official corporate texts. The analysis will focus on whether companies communicate acknowledgement of harm, responsibility acceptance, corrective action, governance-related commitments, and stakeholder engagement, or whether they rely primarily on general value statements, abstract ethical promises, and broad transparency claims. Accordingly, the qualitative research assumes that crisis responses containing specific accountability and reform signals will appear more substantive, while responses dominated by vague ethical discourse will display characteristics of symbolic communication.

Hypotheses for quantitative research:

H1. Crisis responses which include concrete corrective action will be perceived as more trustworthy than responses based mainly on abstract ethical commitments.

The first hypothesis is based on recent studies about trustworthiness and accountability in the tech industry. Keymolen (2024) argues that trustworthy technology firms should be evaluated against three interconnected criteria: assurances, competence, and commitment — meaning that people are more likely to trust a company when its communication demonstrates not only good intentions, but also the practical capacity and willingness to act responsibly (Keymolen, 2024). Similarly, Novelli, Taddeo, and Floridi (2024) define accountability in AI as a relation of answerability that goes beyond vague ethical aspirations and requires verifiable, enforceable follow-through (Novelli et al., 2024). In the context of AI-related ethical scandals, this suggests that responses featuring specific corrective actions, governance reforms, and verifiable follow-up measures are expected to be evaluated more favourably than those based primarily on vague ethical language or abstract pledges. The present study therefore expects crisis responses with clear corrective and governance-related content to be perceived as more trustworthy than responses based primarily on ethical rhetoric.

H2. Crisis responses that include clear accountability signals will be perceived as more accountable than responses limited to general transparency claims.

The second hypothesis draws on research showing that transparency and accountability, while often treated as interchangeable, are functionally distinct in how stakeholders evaluate corporate responses. Schilke and Reimann (2025) describe what they call a transparency dilemma: rather than automatically building trust, disclosure can actually work against a company by signalling to audiences that their autonomy has been constrained — prompting them to question whether the organisation's practices are socially acceptable at all. What matters for positive evaluation, their work suggests, is not the volume of information released but whether the company demonstrates genuine answerability for what occurred (Schilke & Reimann, 2025).

This is consistent with findings by Lim et al. (2025), who show that stakeholders tend to use a company's visible commitment to privacy, security, and safety as a cognitive shortcut when forming broader judgements about its ethical conduct. The hypothesis therefore predicts that crisis responses which include explicit accountability signals — such as direct acknowledgement of fault and verifiable commitments to remedy the situation — will be evaluated as more accountable than responses that rely primarily on general transparency language without connecting it to enforceable action.

H3. Crisis responses that include stakeholder-oriented communication will be perceived as more sincere than responses limited to general transparency claims.

The third hypothesis is grounded in the argument that how a company communicates — specifically, whether it speaks to and about those affected — shapes sincerity perceptions in ways that go beyond the informational content of the response itself. The transparency dilemma identified by Schilke and Reimann (2025) is again relevant here: disclosure that reads as procedural or self-protective can reduce rather than enhance trust, because it disrupts the taken-for-grantedness of a situation and makes audiences more alert to the possibility that the company is managing its image rather than genuinely engaging with the problem. In this context, what closes the sincerity gap is not more information but a different quality of communication — one that acknowledges affected parties directly, addresses their specific concerns, and signals that the company recognises obligations beyond its own reputational interests (Schilke & Reimann, 2025). When that quality is absent, responses tend to read as internally oriented — framing the crisis in terms of what the company did or plans to do, rather than what those harmed experienced or need.

The hypothesis therefore predicts that crisis responses oriented toward the concerns and experiences of affected stakeholders will be rated as more sincere than responses that

remain within the company's own framing of events.

For empirical testing, these research assumptions and hypotheses will be evaluated using the following analytical criteria: perceived trustworthiness, perceived accountability, perceived sincerity, and perceived ethics-washing. These criteria are grounded in the theoretical sources described above and are directly operationalised in the questionnaire items presented in Section 2.5.

2.4 Justification and description of the research methods

For the efficient gathering of data about the crisis communication during AI-related ethical scandals for Big Tech firms, it’s necessary to use a method which can accommodate a larger sample, This section provides a justification and description of the research methodology, which were chosen to test the research hypotheses and research assumptions. This thesis integrates qualitative and quantitative methodologies to achieve two interrelated objectives: firstly, to discern recurring patterns in the framing of AI-related ethical scandals by Big Tech companies; secondly, to evaluate stakeholder interpretations of specific communication elements concerning trustworthiness, legitimacy, accountability, sincerity, corrective-action credibility, and perceived ethics-washing. Case study analysis and online questionnaires were chosen as the primary data-collection techniques, because they are widely used, allow rapid and cost-effective outreach to large audiences, and offer practical advantages such as ease of design, time efficiency, and minimal effort. These methods there-fore provide a suitable and robust foundation for the present investigation.

Comparative case study analysis. Comparative case study analysis serves as the primary qualitative method in this study. It is particularly well suited to a thesis that examines several Big Tech companies — each functioning as a kind of "private governor" and architect of the digital environment — as they navigate similar AI-related ethical controversies, rather than focusing on a single company in isolation (Reid & Ringel, 2025). This approach makes it possible to identify recurring patterns, shared features, and meaningful differences in how firms respond to the "legitimacy gap" when facing comparable forms of public criticism (Wen & Holweg, 2024). Wen and Holweg's (2024) study of four major tech companies that faced public controversy over facial recognition technology supports this directly: they found "very distinct reactions" and identified four main organisational response types to AI ethical failure. This kind of methodology is especially well matched to AI-related scandals specifically. Coombs and Tachkova (2022) describe these controversies as a form of "scansis" — a hybrid of technical crisis and moral scandal — that generates a level of public outrage and reputational damage

well beyond what a standard product failure or service disruption would produce. Understanding how a company navigates that kind of pressure requires attending to the full context of a case rather than examining individual messages in isolation, and case study analysis is designed to do exactly that (Coombs & Tachkova, 2022).

The comparative design also has direct empirical backing. Wen and Holweg (2024) applied a similar case-based approach in their facial recognition study and found that corporate responses to AI ethical failures follow recognisable patterns — Deflection, Improvement, Validation, and Pre-emption — rather than being random or purely situational. Their typology forms the main analytical framework for the qualitative strand of this thesis. Jang and Plaisance (2025) likewise used comparative textual analysis to explore how Big Tech firms differ in their responsible AI communication, revealing meaningful variation in which ethical commitments companies foreground and how they define those commitments (Jang & Plaisance, 2025).

The cases selected for this thesis were chosen deliberately, based on a defined set of criteria. Each case involves a documented AI ethical failure — understood as the deployment of an AI application that contravenes social norms and provokes sustained public condemnation — in areas such as algorithmic bias, surveillance, or privacy violations (Wen & Holweg, 2024). The data for each case are drawn from official company-authored materials, including press releases, newsroom posts, executive statements, and public policy updates issued during or after the controversy (Wen & Holweg, 2024). This focus is methodologically appropriate because the study aims to examine how companies publicly manage the "stage" of their data relationships — constructing narratives of harm and reform through strategic communication in order to maintain or recover corporate legitimacy (Reid & Ringel, 2025).

Comparative case study analysis also plays an important functional role in supporting the quantitative component of the study. By identifying real-world response patterns that recur across cases — such as the use of symbolic ethics principles or the announcement of verifiable corrective actions — the qualitative analysis provides the basis for constructing the survey vignettes (Coombs & Tachkova, 2022). This ensures that the questionnaire stimuli reflect the genuine communication strategies Big Tech firms actually use to address the "AI trust gap," which in turn makes it possible to connect observed organisational behaviour with stakeholder assessments of trustworthiness and accountability (Lim et al., 2025).

There is a further, more practical reason for grounding the quantitative component in case study findings. When questionnaire scenarios are built from how real companies have actually framed their crisis responses, the resulting instrument is measuring something that exists in the world rather than something hypothetical. This strengthens the ecological validity of the questionnaire and helps ensure that the survey measures something that corresponds to

real-world practice.

Online questionnaire method. The quantitative component of this study is based on an online questionnaire — a method in which a survey instrument is distributed electronically and participants record their responses by selecting, ranking, or typing answers that are captured automatically (Wen & Holweg, 2024). The choice of this method is grounded in the tradition of social-mediated crisis communication (SMCC) research, which has long relied on quantitative approaches — particularly surveys and experiments — to study how stakeholders engage with organisations in digitally mediated contexts (Cheng et al., 2022). An online format is especially appropriate for this study, given that it enables the collection of quantitative data from a broad and varied population, which is necessary for examining patterns in perceived governance credibility and trustworthiness (Cheng et al., 2022).

Several specific advantages of the online format are worth noting here. First, it produces a more diverse respondent pool than either face-to-face or telephone approaches, which is important for a study focused on how everyday AI users — rather than specialists or students — evaluate corporate communication (Cheng et al., 2022). Contemporary online research platforms such as Prolific or CloudResearch allow researchers to recruit large numbers of participants from different countries and professional backgrounds (Lim et al., 2025). Recent AI scholarship illustrates what this makes possible: Lim et al. (2025) conducted a national survey of 819 generative AI users, while Yoganathan et al. (2025) ran a multi-country study involving over 10,000 respondents to examine cross-cultural variation in attitudes toward AI safety (Yoganathan et al., 2025).

Second, online questionnaires offer respondents a greater degree of anonymity, which tends to encourage more candid answers on attitudinal questions — something particularly relevant here, since participants are being asked to assess the credibility of some of the world's most prominent companies (De Almeida & Dos Santos Júnior, 2025). The format also allows for features such as automated skip logic and built-in validation checks, which reduce response errors and ease the burden on both the researcher and the participant (De Almeida & Dos Santos Júnior, 2025). Platforms like Prolific have additionally been shown to yield high-quality data with low rates of failure on attention checks, which helps protect the integrity of the measurements collected (Lim et al., 2025).

Third, the online format supports efficient data collection with automatic recording of responses, substantially reducing the risks associated with manual entry and cutting down the time needed for data preparation. Traditional survey methods often involve labour-intensive processes — posting, data entry, cleaning — that introduce delays and additional error. Web-based surveys eliminate most of these steps, allowing for rapid collection, immediate recording,

and the kind of quality controls that would be impractical in paper-based formats. The anonymity of the digital environment also plays a role here: respondents tend to feel more comfortable sharing genuine opinions online than they would in an in-person or telephone setting, which can reduce social desirability bias and produce more accurate attitudinal data overall.

The core of the questionnaire used a vignette-based design. Rather than asking respondents for their general views on AI companies, each participant read a brief, factual account of a real AI-related scandal involving one of the three companies, followed by a standardised summary of that company's official public response. They were then asked to rate the response across five constructs: trustworthiness, accountability, legitimacy, sincerity, and ethics-washing perception. This design is useful because it anchors the evaluation in a specific stimulus — a real corporate communication — rather than in general impressions or prior opinions, which allows for more controlled comparisons across the three company conditions.

Overall, the combination of comparative case analysis and online questionnaire follows a triangulation logic. The qualitative strand examines how companies construct their crisis narratives. The quantitative strand examines how those narratives are received. Neither approach alone would be sufficient to address the research problem, because corporate communication is a two-sided phenomenon: it is both strategically constructed by organisations and socially interpreted by audiences. Bringing both sides into the same empirical study is what makes the mixed-method design appropriate for this thesis.

2.5 Substantiation and description of the data collection and processing methods, sample, and data sources

This section describes the data used in the empirical study, the reasons for the chosen methods of collecting and processing the data, and the sampling strategy used for both research strands in detail. The research employs a mixed-method design, integrating qualitative case study analysis with quantitative data obtained via an online questionnaire. These two methods were meant to work together: the qualitative research looks at how Big Tech firms have communicated during AI-related ethical scandals, when the quantitative captures how stakeholders evaluate different characteristics of such communication.

The selection of data sources, cases, and respondents was guided by the research objectives, which were established in Section 2.1 and the theoretical framework developed in Part 1. Specifically, the qualitative analysis is anchored in published and publicly available

corporate crisis communications, including official statements, newspaper posts, blog publications, and policy updates, produced by selected Big Tech firms in response to documented AI-related situations. The quantitative data were gathered through an online questionnaire targeting general adult users of AI-powered technologies across different countries of Europe.

2.5.1 Qualitative research data collection

The qualitative aspect of this study is grounded in a comparative case analysis of three prominent technology companies - Google, OpenAI, and Amazon - all of which encountered a publicly recognised ethical controversy concerning AI from 2018 to 2024. This design was selected as it facilitates a systematic analysis of the similarities and differences in the crisis communication strategies of companies subjected to analogous ethical pressures, yet responding divergently. This design is especially suitable for the current study, which seeks to discern recurring patterns in corporate crisis communication and assess the efficacy of response strategies.

Four criteria were used to choose the three cases. First, all of the companies are considered "Big Tech" companies, which means they are large, global technology companies with a lot of market power and a lot of work on AI development. Second, each case involves a publicly documented AI ethical failure. This is when an AI application or AI-related organisational practice is used that goes against social norms and gets a lot of criticism from the public. Third, the chosen cases show different types of scandals, such as problems with safety and military use, governance and accountability, algorithmic bias, and surveillance. This variation enhances the analytical breadth of the study while maintaining the comparability of the cases. Fourth, each case has a well-documented public corporate response, which makes it possible to do a systematic content analysis of official crisis communication.

Table 3 provides an overview of the selected cases, their core controversy, the period under analysis, and the response category assigned using the response typology adapted from Wen and Holweg (2024).

Table 3

Overview of selected cases for qualitative analysis

Company	**Scandal Description**	**Period**	**Controversy Type**	**Response Strategy (Wen & Holweg, 2024)**
Google	Project Maven: AI-powered drone	2017–2019	Safety / Military use	Pre-emption

	surveillance contract with the US Department of Defense			
OpenAI	Governance crisis (Altman firing) and copyright controversy over training data	2023–2024	Governance failure / Copyright	Improvement + Validation
Amazon	Rekognition facial recognition system: racial and gender bias documented by ACLU and MIT researchers	2018–2021	Bias / Discrimination	Deflection

Note: : Compiled by the author based on Wen & Holweg (2024) and publicly available corporate communications.

The qualitative data were drawn from two distinct categories of source material. The first consists of official communications published by the companies themselves during and in the aftermath of each scandal — newsroom posts, corporate blog entries, public statements, executive correspondence, updates to AI ethics or policy pages, and formal replies to criticism from regulators or civil society organisations. Where documents were no longer directly accessible, archived versions were retrieved using the Wayback Machine. The second category encompasses contextual secondary sources: peer-reviewed academic studies, civil society reports, and investigative journalism from established outlets. These were used to situate each case in its broader context and to cross-check corporate claims against external accounts, though the primary analytical focus remained consistently on the companies' own official texts.

To make the cases as comparable as possible, the analysis was bounded by a defined timeframe for each: beginning from the point at which the controversy first escalated significantly in public, and ending with the company's most substantial official response or subsequent clarification. This approach enables both the initial reaction and subsequent efforts to justify, reformulate or defend the company’s position to be captured, while avoiding an overly broad or uneven corpus across cases.

The collected documents underwent analysis through directed qualitative content analysis, a systematic approach in which data are categorised according to predetermined classifications based on the theoretical framework. The analytical categories utilised in this study encompass the framing of responsibility attribution, the prevailing response position on the deflection-to-pre-emption spectrum, the characteristics of accountability and reform signals,

the logic of communication channels and sources, and the existence or nonexistence of markers linked to symbolic ethics-oriented communication. Initially, each case was coded independently; subsequently, a cross-case comparative synthesis was performed to discern recurring patterns and significant differences among firms.

2.5.2 Quantitative research data collection

The target sample size was determined prior to data collection using standard power and precision criteria. Assuming a target population of adult AI-tool users across Europe large enough to be treated as effectively infinite, a confidence level of 95%, a conservative response distribution, and a margin of error of ±5%, the minimum required sample size was calculated as follows (see *Formula 1*):

$$n = \frac{z^2 * p(1-p)}{e^2}, \quad (1)$$

where: n — minimum required sample size; z — z-score corresponding to the chosen confidence level (1.96 for 95%); p — estimated population proportion (0.5, conservative estimate); e — acceptable margin of error (0.05).

Substituting the values: $n = (1.96)^2 \cdot 0.5 \cdot 0.5 / (0.05)^2 = 3.8416 \cdot 0.25 / 0.0025 = 385$.

The achieved sample of N = 389 respondents satisfies both the power-based minimum (n = 246) and the precision-based threshold (n = 385), confirming that it is adequate for the planned statistical analyses.

The quantitative part of the study was carried out through an online questionnaire distributed using Google Forms, available between 3 April and 3 May 2026. The questionnaire was shared through social media platforms (Instagram, Telegram, Threads) and via email. Respondents were recruited through convenience and snowball sampling, meaning that participants were reached through the researcher's own networks and were invited to share the questionnaire further.

The questionnaire was structured in three main parts. The first part collected basic demographic information — age, gender, country of residence, and educational level — along with questions about how often participants use AI-powered tools and how familiar they are with AI-related controversies. The second part asked about general attitudes toward Big Tech companies and their ethical AI commitments. The third part was the main evaluation section: participants read a short, factual description of a real AI-related scandal involving one of the three selected companies, followed by a standardised summary of that company's official public

response, and were then asked to rate the response using a five-point Likert scale (1 = Strongly disagree, 5 = Strongly agree).

To ensure comparability across conditions, respondents were randomly assigned to one of three questionnaire versions, each centred on a specific company case: Google, OpenAI, or Amazon. Each respondent evaluated only one case, which prevented survey fatigue and reduced carry-over effects between conditions. The same evaluation items and response format were used across all three versions. The crisis scenarios and response summaries were standardised as closely as possible in terms of length, structure, and tone, in order to minimise presentation-format differences that could affect evaluations independently of the response content itself.

The survey targeted adults aged 18 and over who had used at least one AI-powered tool or service. AI use was defined broadly to include chatbots (e.g., ChatGPT, Gemini, Copilot), voice assistants (e.g., Siri, Alexa, Google Assistant), recommendation systems (e.g., Netflix, Spotify, YouTube), AI image or video generation tools (e.g., Midjourney, DALL-E), and AI writing or productivity tools. This inclusive definition was selected because the study aimed to examine the perceptions of everyday users of AI technologies rather than exclusively technically knowledgeable or expert audiences.

A convenience sampling strategy was applied, which is consistent with the exploratory and comparative nature of the study. Although convenience sampling limits the statistical generalisability of the findings, it is an appropriate approach for a master's thesis that uses an online questionnaire to compare stakeholder evaluations of structured crisis response scenarios. Participants were informed that the study was conducted for academic purposes only, and that all responses were anonymous and participation was voluntary.

A total of 389 respondents completed the questionnaire. Of these, 128 evaluated Google's response to the Project Maven controversy, 130 evaluated OpenAI's response to the governance and copyright controversy, and 131 evaluated Amazon's response to the Rekognition facial recognition controversy. A descriptive overview of the sample is presented in *Table 4*. The full questionnaire instrument is provided in *Annex 3*.

Table 4

Descriptive overview of questionnaire respondents (N = 389)

Variable	**Category**	**N (%)**
Gender	Male	180 (46.3%)
	Female	201 (51.7%)
	Other / Prefer not to say	8 (2.1%)

Age group	18–24	225 (57.8%)
	25–34	126 (32.4%)
	35–44	27 (6.9%)
	45–54	10 (2.6%)
	55+	1 (0.3%)
AI usage frequency	Daily	109 (28.0%)
	Several times a week	179 (46.0%)
	Occasionally	92 (23.7%)
	Rarely or never	9 (2.3%)
Company evaluated	Google (Project Maven)	128 (32.9%)
	OpenAI (Governance / Copyright)	130 (33.4%)
	Amazon (Rekognition)	131 (33.7%)

Note: compiled by the author based on primary questionnaire data.

The main evaluation section of the questionnaire assessed four constructs: trustworthiness, accountability, sincerity, and ethics-washing perception. Each construct was measured using multiple Likert-scale items derived from the theoretical framework developed in Part I of the thesis and adapted to the context of corporate responses to AI-related ethical scandals. This approach ensured internal consistency between the research questions, hypotheses, and the empirical indicators used in the survey instrument.

A descriptive overview of the questionnaire design in *Table 5*.

Table 5

Online questionnaire design: evaluation constructs and theoretical grounding

Construct	**Survey Items**	**Theoretical Source**
Trustworthiness	This company is being honest about what happened. After reading this response, I feel more confident about this company. I believe this company will actually follow through on its promises.	Keymolen (2024)
Accountability	This company clearly takes responsibility for what went wrong. This company explains what will be done to prevent this from happening again. The response includes concrete and verifiable commitments, not just general promises.	Novelli, Taddeo & Floridi (2024)
Sincerity	This response feels genuine rather than scripted. I believe this company truly wants to fix the problem.	Hornsey et al. (2024)
Ethics-washing perception	This response feels like a way to avoid real accountability. This company seems more interested in protecting its image than fixing the problem. This statement feels like it was written mainly for public relations purposes.	Van Maanen (2022); Papyshev & Chan (2024)

Source: compiled by the author based on the theoretical framework established in Part I. All items were rated on a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree). Items measuring ethics-washing were coded so that higher values indicate stronger perceived ethics-washing.

The quantitative data were analysed using descriptive statistics and inferential tests. Mean scores and standard deviations were calculated for each construct across the three company groups. One-way ANOVA was then used to determine whether differences in mean scores across the three case conditions were statistically significant, with Tukey post hoc tests applied where the homogeneity of variances assumption was met and Games-Howell post hoc tests applied where it was violated. Pearson correlation analysis was conducted to examine the relationships between constructs across the full sample. The integration of the qualitative and quantitative strands follows a triangulation logic: the qualitative analysis examines how each company strategically constructs its crisis communication, while the quantitative questionnaire captures how that communication is perceived and evaluated by stakeholders. Together, these two approaches provide a more comprehensive account of crisis communication effectiveness than either method could achieve independently.

2.6 Limitations of the study and evaluation of the reliability of the results

There are several methodological limitations in this study that should be acknowledged clearly. The qualitative part relies on purposive case selection, which means the findings cannot be generalised beyond the three chosen companies. The corporate documents analysed are inherently strategic texts – they were written to manage perception, not to provide a neutral account of events – so the analysis can identify how firms frame responsibility and reform, but it cannot determine whether these claims reflect what actually happened internally. The quantitative part uses a convenience and snowball sample, which is not statistically representative of the broader population. And the vignette design, while useful for controlled comparison, necessarily simplifies the complexity of real crisis situations.

Limitations of qualitative research:

- The analysis of corporate crisis communications relies exclusively on publicly available materials, such as official statements, publications from the press office, and policy documents. Therefore, it cannot account for internal decision-making processes, communications directed at employees or investors, or unpublished responses that may have shaped stakeholder perceptions in ways that are not visible in the public record.

- The three cases - Google's Project Maven, OpenAI's governance and copyright controversies, and Amazon's Rekognition - were selected to represent different types of scandal and response strategy. While this diversity strengthens the comparative dimension of the study, it also limits the extent to which findings can be generalised to make universally applicable conclusions about crisis communication in the tech industry. Each case has its own historical, regulatory and industry-specific context.
- The directed content analysis was conducted solely by the researcher, without independent inter-rater coding. Although the analytical categories were derived from an established theoretical framework (Wen & Holweg, 2024) and applied consistently, the absence of a second coder means that inter-rater reliability cannot be formally established. This is an acknowledged limitation of qualitative case study research conducted at master's thesis level.

Limitations of the quantitative research:

- The sample recruited for the online questionnaire was gathered through convenience sampling and is therefore not statistically representative of the broader population of AI technology users.
- The findings describe the perceptions of the sample group only and should not be generalised to any specific national or demographic group without further research.
- The vignette design offers controlled comparability across three company conditions but requires respondents to form evaluations based on a written scenario rather than their accumulated prior experience with the company. This approach may flatten the nuances of real-world reputation perceptions, which typically develop over time and across multiple touchpoints.
- Respondents' prior awareness of the specific AI scandal described in the vignette was not fully controlled for. Participants who had already formed strong opinions about a given company - whether positive or negative - may have responded to the evaluation items in ways that reflected their pre-existing attitudes rather than providing a fresh assessment of the communication itself. This potential for confirmation bias cannot be fully eliminated in a cross-sectional questionnaire design.
- All four evaluative constructs - trustworthiness, accountability, sincerity and ethics-washing perception - are self-reported attitudinal measures. As with any Likert-based instrument, they capture stated perceptions rather than observed behaviour, and social desirability effects may have influenced how some respondents formulated their answers.

2.7 Evaluation of the reliability of the results

The reliability of the quantitative questionnaire was assessed using Cronbach's Alpha, which evaluates the internal consistency of a set of Likert-scale items measuring the same latent construct. A Cronbach's Alpha value of 0.70 or higher is conventionally considered acceptable for survey instruments in the social sciences (Nunnally, 1978), indicating that the items within each construct are coherently measuring the same underlying dimension. The reliability analysis of the questionnaire constructs is presented in *Table 6*. Detailed item-level reliability statistics for all four constructs are provided in *Annex 4*.

Table 6

Reliability analysis of the questionnaire constructs

Construct	Cronbach's Alpha	N of Items
Trustworthiness	0.799	3
Accountability	0.739	3
Sincerity	0.668	2
Ethics-washing perception	0.703	3

Note: Compiled by the author based on primary questionnaire data analysed in SPSS. Compiled by the author based on primary questionnaire data analysed in jamovi. A Cronbach's Alpha ≥ 0.70 is conventionally considered acceptable (Nunnally, 1978); ≥ 0.65 is acceptable for exploratory research.

The trustworthiness scale demonstrated good reliability ($\alpha = 0.799$), as did the accountability scale ($\alpha = 0.739$) and the ethics-washing perception scale ($\alpha = 0.703$). The sincerity scale produced a coefficient of $\alpha = 0.668$, which falls marginally below the conventional 0.70 threshold. This is acknowledged as a limitation; however, it is considered acceptable in the context of an exploratory master's thesis, particularly given that two-item scales structurally constrain the maximum achievable alpha value (Eisinga, Grotenhuis, & Pelzer, 2013). The items comprising the sincerity scale were theoretically grounded and internally coherent in their face validity. Overall, the reliability analysis confirmed that the questionnaire instruments were sufficiently consistent to justify the use of composite mean scores in the subsequent statistical analysis.

3. RESEARCH RESULTS AND DISCUSSION OF BIG TECH FIRMS' CRISIS COMMUNICATION DURING AI-RELATED ETHICAL SCANDALS

This chapter presents the empirical findings of the study in two sequential strands — qualitative and quantitative — and integrates them into a set of practical recommendations. The qualitative strand examines how Google, OpenAI, and Amazon constructed their official crisis communications during three AI-related ethical scandals between 2017 and 2024. The quantitative strand reports stakeholder evaluations of those communication strategies on four constructs: trustworthiness, accountability, sincerity, and ethics-washing perception. Taken together, the two strands offer a two-sided account of crisis communication effectiveness in the AI ethical scandal context.

3.1 Qualitative research results: Case study analysis

The qualitative analysis examines three Big Tech firms — Google, OpenAI, and Amazon — and how each communicated with external stakeholders during publicly documented AI-related ethical scandals between 2017 and 2024. The selected cases cover a range of ethical controversy types, including military applications of AI, governance breakdown and copyright-related tensions, and algorithmic bias in facial recognition systems. Taken together, they provide a solid empirical basis for examining how major technology companies handle public scrutiny when their AI-related activities attract global criticism, stakeholder pressure, and reputational damage.

The qualitative analysis was conducted using directed content analysis, guided by five coding categories: (1) the framing of responsibility attribution; (2) the position of the response strategy on the deflection-to-pre-emption spectrum; (3) the nature of reform and accountability signals; (4) channel and source logic; and (5) ethics-washing risk indicators. Each case is analysed separately to identify its primary communication patterns.

3.1.1 Case 1: Google - Project Maven (2017–2019)

Background and crisis trigger. Project Maven was a contract signed between Google Cloud and the United States Department of Defense in late 2017, under which Google agreed to apply machine learning techniques to analyse footage captured by military drones. The

project was initially kept confidential within the company (Funmilola et al., 2024). The situation became public in March 2018, when The Intercept and Gizmodo reported on the contract, immediately triggering both an internal and external crisis. Over 4,000 Google employees signed an open letter to CEO Sundar Pichai demanding that the contract be cancelled, and approximately twelve employees resigned in protest (Wen & Holweg, 2024). The core ethical objection was not that the AI system was autonomous or directly lethal — Google maintained that it was neither — but that the technology supported a military targeting infrastructure in ways that employees and civil society groups saw as fundamentally incompatible with the company's stated values and its unofficial founding principle of "Don't be evil" (Funmilola et al., 2024).

In its initial public communications, Google adopted a partial responsibility framing: it acknowledged the contract's existence but tried to minimise its significance, describing the software as "non-offensive" and framing it as a tool for saving lives (Wen & Holweg, 2024). Internally, executives characterised the project as a limited $9 million initiative with a clearly defined scope, apparently hoping this framing would reduce the perceived gravity of the company's involvement (Funmilola et al., 2024). This approach became difficult to maintain when subsequent leaks revealed that senior leadership had privately viewed Project Maven as a potential "on-ramp" to a $250 million annual defence contract — a disclosure that substantially damaged employee trust (Funmilola et al., 2024).

Responsibility attribution fraiming: pre-emption. What stands out about Google's response is that the company did not try to blame the controversy on external misunderstandings or third parties. Instead, it accepted that the contract had created genuine ethical tensions within its own workforce — a more inward-facing framing of responsibility than is typically seen in comparable industry cases. Google ultimately moved toward a pre-emption strategy, announcing in June 2018 that it would not renew the contract and subsequently publishing a new set of AI principles that explicitly prohibited the development of AI for weapons use (Funmilola et al., 2024).

This two-pronged response — contract withdrawal plus published ethics principles — constitutes a substantive pre-emption because it combined a concrete material action (non-renewal) with a governance document that formally codified the withdrawal rationale and attempted to prevent analogous controversies in future. This is consistent with the distinction drawn in Section 1.4.3 between rhetorical reform signals and structural reform signals: the AI Principles document was not merely a statement of regret but an operationalised policy commitment that the company has since been held publicly accountable to.

Reform and accountability signals. Google’s reform efforts followed both procedural

and substantive paths. Procedurally, the company acknowledged a governance vacuum and established internal review programmes, such as its Responsible Innovation team, to evaluate the legal and ethical implications of future contracts. On the substantive side, the AI Principles document explicitly prohibited four categories of application:

- Weapons systems.
- Surveillance technologies that violate international norms.
- Technologies that contravene international law.
- Tools that work against human rights.

However, subsequent events weakened the credibility of these signals. In March 2019, Google's Senior Vice President for Global Affairs confirmed in an internal email that an unnamed third-party company would take over the Project Maven workload using Google Cloud Platform infrastructure. Although the email maintained that this arrangement remained consistent with Google's AI Principles, civil liberties observers and former employees read it as a structural continuation of the military relationship under a different contractual label (The Intercept, 2019). This episode points to a broader vulnerability in pre-emption as a crisis response: even a substantive withdrawal can come to look symbolic if the underlying commercial relationship is simply reorganised rather than terminated.

Channel and source logic. Google's crisis communications operated across three main channels. Internal communications — all-hands meetings, executive emails, and the internal Maven Conscientious Objectors forum — preceded and shaped the external response. External corporate communications were published through Google's official blog and via CEO Sundar Pichai's public statements, including his keynote address at Google I/O in May 2018, where he acknowledged that navigating new AI territory required proceeding "carefully and deliberately." Media coverage was predominantly reactive: Google never issued a press release about the contract itself, and all major public disclosures came from investigative journalism rather than voluntary transparency. This reactive information environment constrained Google's ability to shape the crisis narrative — a structural disadvantage it shares with the other two cases in this analysis.

Ethics-washing risk assessment. The ethics-washing risk associated with Google's Project Maven response is assessed as low-to-moderate. The primary reason for this relatively favourable assessment is that the most substantive crisis communication action — non-renewal of the contract — was a decision with real commercial consequences, not merely a discursive commitment. The AI Principles document, though published under crisis conditions, has remained operative and has been cited by Google in subsequent public controversies as a binding normative framework. These features set Google's response apart from a purely

performative ethics exercise.

Two factors push the assessment from low toward moderate:

1. First, *Structural Continuation*: the infrastructure transfer to a third-party contractor suggested that the underlying commercial interest in defence work had not been fully abandoned.
2. Second, *Narrow Scope of Principles:* including cloud computing contracts that did not technically fall within the AI Principles' weapons exclusion — indicated that the principles' scope was narrower than the initial announcement had implied (MIT Technology Review, 2021).

The tension between the principled public position and the continued commercial relationship with defence clients is where the primary ethics-washing risk in this case lies.

3.1.2 Case 2: OpenAI – Governance Crisis and Copyright Controversy (2023–2024)

Background and trigger of the crisis. OpenAI's crisis actually involved two separate but connected controversies happening around the same time. The first one was a governance crisis that started on 17 November 2023, when the non-profit board suddenly fired CEO Sam Altman. The reason given was that he had not been honest enough with the board, which made it impossible for them to properly do their job of overseeing the company. The board existed specifically to make sure that OpenAI's public benefit mission came before profit or investor interests. According to reports, Altman's lack of transparency was described as "completely unworkable" — especially regarding his ownership of the OpenAI startup fund and the launch of ChatGPT, which the board had apparently only found out about through Twitter. What happened next was almost immediate: 702 out of 770 employees said they would quit if Altman was not brought back. Microsoft pushed hard for his reinstatement, and five days later he was back as CEO (Trautman, 2024).

The second controversy concerned the large-scale use of third-party intellectual property in training models such as ChatGPT and the Sora video generator. Critics and artists argued that OpenAI's operating model had increasingly prioritised aggressive monetisation and the rapid release of consumer products over the company's original altruistic mandate. This tension was further sharpened by allegations that the company had exploited beta testers and artists for unpaid labour under the pretence of collaborative development (Lim et al., 2025).

This analysis treats the two controversies together, since they emerged during the same crisis period and were handled by OpenAI's crisis communications in overlapping ways. Both ultimately pointed to the same structural tension at the heart of the OpenAI situation: the

company's founding non-profit mission of developing AI "for the benefit of all humanity" was increasingly difficult to reconcile with a growing commercial model that included a multi-billion dollar Microsoft investment, rapid consumer product launches, and aggressive monetisation of intellectual property produced in part by uncredited creators.

Responsibility attribution. The way OpenAI framed responsibility changed as the crisis developed. At first, the board accused Altman of personal misconduct. Later, after he came back, board chair Bret Taylor presented the results of an internal review and concluded that the previous board had acted in good faith but had simply not anticipated how much instability their decision would cause. This shift in framing was significant — it moved the focus away from specific ethical failures, like misrepresenting safety processes, and toward a more abstract story about organisational incompatibility between the non-profit board and the commercial side of the business. Unlike Google's response to the Maven controversy, which directly acknowledged that internal ethical concerns were legitimate, OpenAI's approach was mostly about getting back to normal operations as quickly as possible and protecting its market position.

Response strategy: structural reorganisation and validation. OpenAI's main response was to restructure its leadership in a way that would look credible to commercial stakeholders. The fired board members were replaced by well-known Silicon Valley figures, including Bret Taylor and Larry Summers — a move that was clearly intended to signal institutional maturity and alignment with the market. This fits the validation strategy from the Wen and Holweg (2024) typology, where a company uses association with respected external figures to rebuild legitimacy. At the same time, OpenAI expanded its red-teaming and safety frameworks, presenting this as evidence of a continued commitment to responsible development. Former employees, though, have said that safety work was increasingly pushed aside as the company rushed to stay competitive.

Reform and accountability signals. Most of OpenAI's reform signals were met with scepticism, and not without reason. Replacing the board was a real change, but many people read it as a shift toward more commercial control rather than stronger ethical oversight. The company's public commitment to developing AGI that "benefits all of humanity" lost credibility when, at the same time, it dissolved the internal team working on long-term AI safety research. Then came the "Right to Warn" open letter, signed by current and former employees, which alleged that OpenAI had used confidentiality and non-disparagement agreements to stop people from raising ethical concerns internally. That directly contradicted everything the company had been saying publicly about transparency.

Several specific factors eroded the credibility of these signals further:

- **Commercial prioritisation:** The speed of Altman's reinstatement — just five days — combined with the subsequent departure of safety-focused board members including Helen Toner and Tasha McCauley, strongly suggested that commercial continuity had been prioritised over the mission-level concerns that originally triggered the crisis.
- **Safety gaps:** Former board member Helen Toner later disclosed that the board's concerns had extended beyond Altman's candour to fundamental questions about whether OpenAI's safety practices were keeping pace with its commercialisation. These concerns were not substantively addressed by the governance outcome.
- **Suppression of dissent:** The "Right to Warn" open letter alleged that the company had deployed restrictive confidentiality and non-disparagement clauses to prevent employees from raising ethical concerns, directly contradicting its public claims of openness and accountability.
- **Weak copyright reforms:** In the copyright domain, proposed remedies have been widely viewed as weak or entirely symbolic. OpenAI's Sora model has been described by critics as an example of "participation washing," with artists alleging they were drawn into unpaid labour under the guise of collaboration. The company has not disclosed its training data or proposed systemic licensing arrangements for independent creators, leaving its commitments to structural reform in this area largely unsubstantiated..

Channel and source logic. OpenAI's communication approach differed markedly between the governance crisis and the copyright controversy, with the two situations handled through quite different platforms and registersThe governance crisis was very public and fast-moving — it played out through official website statements, posts on X (formerly Twitter), and internal messages that quickly became public. Because everything was happening on Twitter at the same time — employees, investors, and board members all posting simultaneously — the narrative was almost impossible to control. Altman himself was very active on the platform, posting that he "loved OpenAI" and eventually confirming his return as CEO directly through his own account, which completely bypassed any formal communication strategy and made the whole thing move faster than it might have otherwise.

The copyright controversy was handled very differently. Here, the company relied mostly on legal filings, official blog posts, and responses to journalists rather than proactive communication. There were fewer voluntary disclosures, more legal language, and a general sense that the company was reacting rather than leading. This contrast between the two controversies suggests that OpenAI's approach to crisis communication is not consistent — it adapts based on what kind of crisis it is facing and how much legal exposure is involved.

Analysis of the approach. The difference between these two methods shows that OpenAI's crisis communication isn't the same for all types of controversy. The governance crisis used the openness and interactivity of social media to settle a power struggle within the company, while the copyright issue used a more defensive and procedural logic to deal with long-term legal liability. This inconsistency shows that the level of transparency and control an organization can have over its public story depends on the channel it chooses, whether it's a real-time social platform or a formal legal brief.

Ethics-washing risk assessment. The ethics-washing risk for OpenAI is rated as moderate to high. The core problem is that the company's ethical language — its non-profit structure, its mission statements, its safety commitments — increasingly looks like it is covering for a business model that operates like any other Silicon Valley company: move fast, grow aggressively, and deal with the consequences later. The removal of safety-focused researchers, the handling of the Sora controversy, and the use of legal agreements to silence internal critics all point in the same direction. As long as OpenAI continues pursuing large commercial investments and aggressive IP monetisation while presenting itself as a mission-driven organisation, the gap between what it says and what it does will remain a serious reputational and ethical problem.

3.1.3 Case 3: Amazon – Rekognition Facial Recognition Bias (2018–2021)

Background and crisis. Amazon's Rekognition crisis is one of the clearest examples of what happens when a company deploys a technically flawed AI system in high-stakes settings and then refuses to take responsibility for it. Amazon Web Services had been selling its facial recognition service, Rekognition, to law enforcement agencies for surveillance purposes. The problems had been building for a while, but the crisis really became public in 2018, when researchers and civil society organisations started raising serious questions about whether the technology was reliable enough for real-world use (Wen & Holweg, 2024). In May 2018, the ACLU revealed that Amazon was actively selling the tool to police departments to help identify suspects — which immediately raised questions about accuracy, civil rights, racial discrimination, and what it meant for AI-driven surveillance to become normalised in law enforcement (Wen & Holweg, 2024).

Things got significantly worse in July 2018 when the ACLU published test results showing that Rekognition had wrongly matched 28 sitting members of the U.S. Congress to criminal mugshots. What made this finding especially damaging was that the incorrect matches were disproportionately of people of colour — a group that was significantly underrepresented

in Congress but overrepresented among the false identifications (Wen & Holweg, 2024). At that point, the scandal stopped being just about technical limitations and became a story about a system that was actively reproducing and amplifying racial bias in a context where the consequences could be very serious (Wen & Holweg, 2024).

In January 2019, the academic weight behind these concerns increased significantly when Joy Buolamwini and Timnit Gebru published research showing that commercial facial recognition systems — including Rekognition — performed much worse on dark-skinned women than on light-skinned men. This was not anecdotal criticism anymore; it was peer-reviewed evidence. So by this point, Amazon was facing a crisis on two fronts: not only had it deployed a biased system, but it had kept selling it to law enforcement even after multiple credible warnings (Wen & Holweg, 2024). That combination is what made Rekognition one of the most frequently discussed cases of AI ethical failure in the Big Tech context.

Responsibility attribution. Amazon's position throughout the crisis was consistently defensive. The company never acknowledged that its technology had fundamental flaws. Instead, executives blamed the problems on customers using the tool incorrectly and on methodological errors made by the researchers who tested it. The argument Amazon kept making was that the technology itself was not biased or intrusive — it just depended on how it was used. When it came to the MIT research specifically, Amazon publicly called it "misleading" and claimed the researchers had not followed the right technical protocols. The overall effect of this framing was to make the problem someone else's responsibility — the users, the researchers, the regulators — rather than Amazon's.

Response strategy: deflection and temporary moratorium. In the Wen and Holweg (2024) typology, Amazon's response falls clearly into the deflection category — the most defensive option available. Throughout the crisis, Amazon refused to acknowledge the error rates documented by the ACLU and MIT, and instead kept arguing that the problems were the result of misuse or flawed testing methodology. One specific example was the claim that critics had used an 80% confidence threshold when Amazon recommended 99% for law enforcement applications — a technical argument that largely sidestepped the broader civil rights concerns being raised.

This approach was not just about public statements — it also shaped what the company actually did operationally. Even as the pressure from civil rights organisations, employees, academics, and shareholders kept building, Amazon continued to sell Rekognition to government agencies. This is a meaningful contrast to how Google handled the Maven case, where the company chose to withdraw voluntarily and put new ethical guardrails in place before anyone forced it to. Amazon instead pushed for federal legislation on facial recognition, which

most observers saw as a way to pass the responsibility for regulation to the government rather than taking any themselves. The moratorium on police use of Rekognition only came in June 2020 — and only after the George Floyd protests and the Black Lives Matter movement made the existing position politically untenable. It was later extended indefinitely in 2021.

Reform and accountability. The reform signals from Amazon are seen as the weakest of the three cases being looked at. In 2019, the company gave law enforcement clients guidelines on how to use the software, suggesting that a human should review it and that high confidence thresholds should be used. However, these guidelines were not required and were only suggestions. Later, the ACLU said that at least one police client had used Rekognition without a confidence threshold. This suggests that these rules were more about limiting liability than making sure people were really responsible.

Amazon's most important change is still the moratorium on police use of the facial comparison feature, which was announced in June 2020 and extended indefinitely in 2021. This promise, on the other hand, is only partially fulfilled because it doesn't cover the wider use of the technology in business or its use in immigration enforcement and private security. Also, the moratorium's credibility as a way to hold people accountable has been hurt by reports that the FBI started a project using Rekognition after the moratorium was already in place. This evidence of compliance failure after the fact points to weak internal enforcement.

Channel and source logic. Amazon's crisis communications were reactive and counter-argumentative, primarily utilising corporate platforms such as the AWS and Amazon Web Services Machine Learning Blog to issue posts rebutting individual academic studies. Unlike its competitors, Amazon did not use these channels to address civil rights concerns or acknowledge public interest in algorithmic accuracy; instead, it used them to dispute specific methodological findings.

Researchers Raji and Buolamwini (2024) directly criticised this adversarial posture, noting that Amazon's communications were primarily defensive rather than accountable. Throughout the primary crisis period, Amazon notably avoided:

- Public forums or transparency reports.
- External bias audits or independent evaluations.
- NIST participation, which Microsoft and IBM utilised to verify their internal accuracy claims.

As Amazon's technology performance was entirely self-reported, its communication strategy was incapable of resolving the evidentiary disputes that fuelled the crisis. This approach reinforces the company's classification as employing a deflection strategy, whereby responsibility is consistently attributed to external factors, such as researcher error or customer

misuse, rather than to internal technological flaws.

Ethics-washing risk assessment. The ethics-washing risk for Amazon is assessed as high. The most visible problem is the gap between what the company said publicly — especially during the 2020 Black Lives Matter movement — and what it was actually doing at the same time. Amazon expressed public support for racial justice while maintaining active contracts with police departments that were using Rekognition for exactly the kind of racially biased surveillance that was being protested. Critics were not wrong to call this hypocritical. The company's consistent framing of bias as a misuse problem, rather than a design and deployment problem of its own making, suggests that its ethical communications were primarily intended to protect a profitable product line rather than to address the real harms that had been documented.

3.1.4 Cross-case comparison and qualitative discussion

The three analysed cases revealed both major difference in the crisis communication strategies adopted by each company and a set of patterns which appear to be generalisable to the Big Tech AI ethical scandal context more broadly. *Table 7* summarises the main dimensions of each case across the five coding categories.

Table 7

Cross-case comparison of crisis communication responses

Dimension	Google (Maven)	OpenAI (Governance / Copyright)	Amazon (Rekognition)
Responsibility attribution	Partial — minimised scope, inward-facing	Diffuse — structural, not individual (governance); legal framing (copyright)	Externalising — researcher error, user misuse
Response strategy	Pre-emption	Improvement + Validation	Deflection
Reform signal type	Structural (AI Principles + contract non-renewal)	Mixed: structural (board reform) + rhetorical (copyright)	Rhetorical (guidelines, moratorium framing)
Reform signal credibility	Moderate — eroded by infrastructure transfer	Moderate — eroded by reinstatement outcome	Low — contradicted by commercial behaviour
Channel posture	Reactive (disclosure by media); proactive ethics document	Mixed: reactive (governance); reactive (copyright)	Reactive and counter-argumentative

Ethics-washing risk	Low–Moderate	Moderate–High	Very High

Note: compiled by the author based on directed content analysis of corporate communications and secondary academic sources.

This analysis shows a number of things that can be compared. First, the link between response strategy position and ethics-washing risk is not a straight line. Google's pre-emptive response had a lower ethics-washing risk than Amazon's deflection response, which fits with what Wen and Holweg (2024) predicted. OpenAI's response of improvement and validation, which is theoretically a more flexible strategy than deflection, did, however, pose a moderate to high risk of ethics-washing. This was because the language used to talk about the reform and the result of the governance reform were at odds with each other. The reform made commercial priorities stronger instead of weaker. This means that the risk of ethics-washing in a crisis communication response depends on how well the content of the communication matches the company's actual behaviour, not just on where the strategy falls on the accommodativeness spectrum.

Second, all three cases have a reactive information environment as a structural part of the crisis. In none of the cases did the company voluntarily share the controversy that caused the crisis. Instead, all three were first made public by investigative journalism (Maven), academic research (Rekognition), or internal whistleblowers (OpenAI governance). This pattern connects directly to the observation in Section 1.3 regarding a transparency paradox in Big Tech crisis communication: companies that publish comprehensive voluntary governance documents — AI principles, safety commitments, responsible use policies — simultaneously resist proactive disclosure when specific contentious decisions come under scrutiny. The gap between the broad transparency of ethics documents and the reactive opacity of specific crisis disclosures is a consistent feature of the communication landscape across all three cases.

Thirdly, the reform signals that were most consistently credible across the three cases were those that combined a material action with a governance document. The most obvious example of this pattern is that Google decided not to renew Project Maven when it released the AI Principles. In contrast, Amazon's moratorium was a real action (stopping sales to law enforcement) without a real governance document that acknowledged the evidence of bias that came before it. OpenAI's governance reform produced a new governance document, in the form of board restructuring, but the actual outcome — Altman's reinstatement — ran directly counter to the purpose that restructuring was supposed to serve. This comparison indicates that, regarding ethical scandals in AI, the reliability of reform signals relies on the consistency

between tangible actions and verbal commitments, rather than on either component independently.

3.2 Quantitative research results: Online questionnaire findings

The quantitative strand of this study examined how stakeholders evaluate crisis communication responses across three case conditions. The sample composition and descriptive statistics are presented in Section 2.5.2 and 2.5.3 respectively. The following analysis proceeds in three steps: differences in mean scores across case conditions are examined using one-way ANOVA with post hoc tests, followed by hypothesis testing, and a Pearson correlation analysis of relationships between the four evaluative constructs. Full descriptive statistics and ANOVA output tables are provided in *Annex 6*.

3.2.1 Differences across case conditions

Internal consistency was confirmed prior to analysis (Cronbach's Alpha reported in Table 6, Section 2.6): trustworthiness $\alpha = 0.799$, accountability $\alpha = 0.739$, ethics-washing perception $\alpha = 0.703$, and sincerity $\alpha = 0.668$ — the last considered acceptable for a two-item exploratory scale.

Table 8 presents the mean scores and standard deviations for the four evaluated constructs across the three case conditions.

Table 8

Mean scores and standard deviations by construct and company condition

Construct	Google M	Google SD	OpenAI M	OpenAI SD	Amazon M	Amazon SD
Trustworthiness	3.24	0.654	3.79	0.752	2.02	0.708
Accountability	3.17	0.695	3.80	0.735	2.13	0.643
Sincerity	3.10	0.811	3.65	0.935	2.31	0.858
Ethics-washing perception (reverse-coded)	3.18	0.663	3.74	0.797	2.84	1.012

Note: M = Mean; SD = Standard Deviation. Scale: 1 = Strongly disagree, 5 = Strongly agree. Higher ethics-washing scores indicate stronger perception of symbolic or image-protective communication. Source: Compiled by the author based on primary questionnaire data analysed in jamovi.

The results reveal consistent variation across conditions. OpenAI (Case 2) received the highest scores on trustworthiness (M = 3.79), accountability (M = 3.80), and sincerity (M = 3.65); Amazon (Case 3) received the lowest on all three. Overall sample means clustered near the scale midpoint (Trust = 3.01, Acc = 3.03, Sinc = 3.02, Ethics = 3.25), indicating that respondents were not polarised on average, but case-level differences were substantively large.

One-way ANOVA confirmed statistically significant differences across all four constructs (all $p < .001$). For trustworthiness, $F(2, 386) = 215$; accountability, $F(2, 386) = 195$; sincerity, $F(2, 386) = 78.8$. Tukey post hoc comparisons confirmed that all three pairwise differences were significant for each construct, with mean differences frequently exceeding one full scale point. For ethics-washing perception, Levene's test indicated violated homogeneity ($F(2, 386) = 18.6$, $p < .001$), so Welch's ANOVA and Games-Howell post hoc tests were applied: Welch's $F(2, 252) = 35.7$, $p < .001$, with all pairwise differences significant.

Notably, the ethics-washing pattern ran in the opposite direction to the three positive constructs: Case 2 received both the highest positive evaluations and the highest ethics-washing score (M = 3.74), while Case 3 received the lowest (M = 2.84). This indicates that stakeholders do not evaluate crisis responses on a single good-to-bad dimension — a response can be seen as credible and strategically self-protective simultaneously. The high variance for Case 3 (SD = 1.012) also suggests respondents were more divided about Amazon's response than the others.

3.2.2 Hypothesis testing

The quantitative part of the study was guided by three hypotheses concerning stakeholder evaluations of crisis communication responses. Each hypothesis is assessed below against the ANOVA and post hoc results reported in Section 3.2.1.

3.2.2 Hypothesis testing

The quantitative part of the study was guided by three hypotheses concerning stakeholder evaluations of crisis communication responses, each assessed against the ANOVA and post hoc results reported in Section 3.2.1.

H1 predicted that crisis responses including concrete corrective action would be perceived as more trustworthy than responses based mainly on abstract ethical commitments. H1 is supported. Trustworthiness scores differed significantly across the three case conditions, $F(2, 386) = 215$, $p < .001$, and the ordering of means is consistent with the hypothesis: Case 2 (OpenAI, M = 3.79) scored highest, followed by Case 1 (Google, M = 3.24) and Case 3 (Amazon, M = 2.02). All pairwise post hoc differences were statistically significant, and the

mean difference between Case 2 and Case 3 (MD = 1.77) constitutes a large effect by conventional standards. These results are consistent with Keymolen's (2024) argument that trustworthiness depends on the demonstrated combination of assurances, competence, and commitment: responses presenting observable corrective action were evaluated as substantially more trustworthy than those offering only rhetorical reassurance.

H1 nonetheless requires partial qualification. The finding that Case 2 outscores Case 1 on trustworthiness — despite the qualitative analysis identifying OpenAI's governance outcome as commercially captured — suggests that perceived concreteness is shaped not only by the substantive depth of corrective action, but also by the visibility and elaborateness of the communicative act itself.

H2 predicted that crisis responses including clear accountability signals would be perceived as more accountable than responses limited to general transparency claims. H2 is supported. Accountability scores differed significantly across conditions, $F(2, 386) = 195$, $p < .001$, replicating the trustworthiness hierarchy: Case 2 ($M = 3.80$), Case 1 ($M = 3.17$), Case 3 ($M = 2.13$). Of all four constructs, accountability produced the most consistently hierarchical pattern, with every case meaningfully distinguishable from every other. This aligns with Schilke and Reimann's (2025) transparency dilemma argument: it is not the volume of information disclosed but evidence of genuine answerability — acknowledgement of fault, verifiable remedial commitments, observable governance change — that drives favourable evaluations. Amazon's response, which relied on general transparency language while never acknowledging documented bias, scored more than one full scale point below both other conditions.

H3 predicted that stakeholder-oriented communication would be perceived as more sincere than responses limited to general transparency claims. H3 is partially supported. Sincerity scores varied significantly across conditions, $F(2, 386) = 78.8$, $p < .001$, in the predicted direction: Case 2 ($M = 3.65$), Case 1 ($M = 3.10$), Case 3 ($M = 2.31$), with all pairwise differences statistically significant. However, a methodological caveat applies: the three vignettes differed simultaneously across multiple dimensions, including response strategy type, corrective action depth, and accountability signal strength. It is therefore not possible to attribute the sincerity differences specifically to stakeholder orientation in isolation. What the data confirm is the directional prediction; isolating the effect of stakeholder orientation would require an experimental design in which it is manipulated as a single variable.

Ethics-washing perception as a cross-cutting qualification. The most analytically consequential finding concerns the relationship between positive evaluations and ethics-washing perception. Although Case 2 received the most favourable scores on all three positive

constructs, it also received the highest ethics-washing score (M = 3.74), with Case 1 intermediate (M = 3.18) and Case 3 lowest (M = 2.84). This pattern — running in the opposite direction to trustworthiness, accountability, and sincerity — is not a contradiction but a substantively important finding: stakeholders do not evaluate crisis responses along a single credibility dimension. The positive correlations between ethics-washing perception and all three favourable constructs — trustworthiness ($r = .373$), accountability ($r = .419$), sincerity ($r = .313$), all $p < .001$ — confirm this across the full sample. Responses that communicated more elaborately about ethical commitments invited both stronger positive evaluations and heightened scepticism about genuineness simultaneously, suggesting that investment in visible, structured crisis communication does not eliminate, and may in fact heighten, the perception that it also serves strategic image-management purposes. The full Pearson correlation matrix is presented in *Annex 5*.

3.3 Integrated discussion and practical recommendations

The qualitative and quantitative strands of this study produce findings that neither could generate independently. The case analyses established how Google, OpenAI, and Amazon constructed their crisis narratives — in terms of responsibility framing, reform signals, and ethics-washing risk. The stakeholder questionnaire revealed how those narratives were evaluated on trustworthiness, accountability, sincerity, and ethics-washing perception. This section integrates the most consequential observations across both strands.

3.3.1 The accommodativeness-credibility gap

The central empirical finding of this study is that the formal accommodativeness of a crisis response strategy does not reliably predict how credible that response is perceived to be. This gap between strategic position and stakeholder evaluation constitutes the primary theoretical contribution of the empirical analysis.

Within the Wen and Holweg (2024) typology, the three cases occupy distinct positions on the accommodativeness spectrum: Amazon's deflection is the least accommodative, Google's pre-emption the most, and OpenAI's improvement-and-validation strategy occupies an intermediate position. If accommodativeness translated directly into perceived credibility, the expected ranking would be OpenAI ≥ Google > Amazon. The quantitative data partially confirm this — Amazon's deflection strategy produced the lowest scores across all three

positive constructs, consistent with SCCT predictions that under-accommodative responses generate the highest reputational damage (Coombs, 2023). This part of the pattern holds.

The more analytically important finding concerns the relationship between Google and OpenAI. As reported in Section 3.2, OpenAI consistently outscored Google on all three positive constructs — by statistically significant margins on trustworthiness, accountability, and sincerity — despite the qualitative analysis identifying Google's pre-emptive contract withdrawal as the more substantively credible response, and OpenAI's governance outcome as commercially captured. A company whose governance reform reinstated the CEO whose conduct had triggered the crisis, replaced safety-focused board members with commercially oriented ones, and dissolved its long-term AI risk team nonetheless received higher stakeholder evaluations than the company that voluntarily withdrew from a controversial military contract before external pressure required it.

This pattern suggests that the mechanism driving perceived credibility in AI crisis communication is not accommodativeness per se, but rather the visibility and structural elaborateness of the communicative act. OpenAI's governance crisis unfolded in a highly public, real-time environment — executive statements on social media, rapid board-level communications, and a five-day reinstatement drama that sustained global media attention. This communicative intensity appears to have generated higher positive evaluations not because the governance outcome was substantively responsible, but because the response felt active, institutionally significant, and structurally organised. Stakeholders appear to interpret elaborate, multi-channel governance communication as evidence of accountability regardless of whether the underlying decisions served the organisation's stated mission — a dynamic consistent with Huang and Krafft's (2024) concept of performative platform governance, in which front-stage communicative activity generates positive evaluation even when back-stage institutional decisions move in a contradictory direction.

The accommodativeness-credibility gap therefore has a direct practical corollary: a well-communicated but commercially compromised response can outperform a more genuinely accountable response delivered reactively or without sufficient narrative clarity. This creates a structural incentive to invest in visible governance theatre rather than substantive reform — precisely the ethics-washing dynamic the theoretical framework identified as the central credibility threat in AI ethical scandal contexts. The ethics-washing data make this tension legible: Case 2 received both the strongest positive evaluations and the highest ethics-washing perception scores, confirming that the communicative features that build credibility also heighten suspicion about whether that credibility is earned.

3.3.2 The transparency paradox and reactive disclosure

A second structural pattern cuts across all three cases: in none of them did the company disclose the controversy proactively. Project Maven was exposed by investigative journalism; the Rekognition bias findings came from academic and civil society research; OpenAI's governance crisis was triggered by internal whistleblowers and employee revolt. All three companies entered the crisis communication dynamic already positioned as parties whose behaviour had been exposed rather than voluntarily disclosed.

This matters theoretically. Schilke and Reimann (2025) describe a transparency dilemma in which reactive disclosure — disclosure forced by external investigation rather than internally initiated — signals to audiences that the company would not have disclosed voluntarily. Post-crisis communications therefore begin from a credibility deficit that is structurally difficult to overcome regardless of their content or quality, because the act of communicating is already framed by an implicit admission of prior concealment.

Google's response represents a partial exception. Although the initial Maven disclosure was reactive, Google subsequently withdrew from the contract and published its AI Principles before the contract formally expired — introducing a degree of proactivity into a crisis that had begun defensively. The quantitative data suggest this shift carried evaluative consequences: Google's scores on trustworthiness and accountability, while lower than OpenAI's, substantially exceeded Amazon's, whose response remained reactive and counter-argumentative throughout. Even partial proactivity, introduced mid-crisis, appears to carry a measurable credibility benefit over sustained defensiveness.

The practical implication extends beyond response strategy into governance design. Companies that proactively publish AI deployment decisions, commission independent bias audits before deployment, and maintain genuinely transparent governance processes enter any public scrutiny dynamic from a fundamentally different credibility baseline than those whose practices are first revealed by journalists or civil society researchers. Pre-crisis transparency is not only an ethical commitment — it is a strategic communication asset that determines the conditions under which any post-crisis communication will be received.

This also clarifies the ethics-washing finding. OpenAI received the highest ethics-washing perception scores alongside the most favourable evaluations across the three positive constructs. This coexistence is consistent with respondents recognising that OpenAI's crisis communication was elaborate and structurally active while simultaneously sensing that its transparency was reactive and commercially motivated. The positive correlations between ethics-washing perception and all three favourable constructs confirm this: the same features

that make a crisis response appear credible — visibility, structural elaborateness, explicit ethical framing — also make it more legible as a strategic reputational exercise. Companies that communicate extensively about their ethical commitments under crisis conditions may paradoxically attract more scrutiny about whether those commitments are genuine, not less.

3.3.3 Practical recommendations

The empirical findings of this study — both the qualitative patterns identified across the three case studies and the quantitative results from the stakeholder questionnaire — converge on a set of actionable implications for Big Tech firms facing AI-related ethical scandals. The five recommendations below are grounded directly in the evidence: each is traceable to a specific finding from the case analysis, the ANOVA results, or the integrated discussion, rather than derived from general normative principles. Taken together, they constitute a practical framework oriented toward the core insight of this study — that credibility in AI crisis communication depends less on the formal strategy type deployed and more on the coherence between communicative commitments and operational behaviour, the proactivity of pre-crisis transparency practices, and the specificity and structural credibility of reform signals.

Recommendation 1: Align reform signals with operational actions before communicating them publicly. The most significant credibility risk identified across all three cases arose when companies communicated a commitment to change while their concurrent operational behaviour remained unchanged or moved in a contradictory direction. OpenAI publicly committed to safety-first governance while simultaneously reinstating the CEO whose lack of candour had triggered the board's original decision, replacing safety-focused board members with commercially oriented ones, and disbanding the team responsible for long-term AI risk research. Google published AI Principles explicitly prohibiting weapons-related AI development while internally facilitating the transfer of its Project Maven workload to a third-party contractor via its own cloud infrastructure. In both cases, the gap between communicative commitment and operational behaviour became the primary driver of ethics-washing perception — a pattern the quantitative data confirmed, with Case 2 receiving the highest ethics-washing score ($M = 3.74$) despite simultaneously receiving the most favourable evaluations on trustworthiness, accountability, and sincerity.

Big Tech firms should therefore ensure that announced reforms are operationally implemented — or at minimum publicly scheduled with measurable, time-bound milestones — before they are communicated as crisis responses. Reform signals that precede the operational changes they describe invite audiences to evaluate the gap between the two, and that evaluation

will almost always be unfavourable. If reform commitments cannot be immediately operationalised, they should be framed as forward commitments with specific accountability mechanisms attached — named responsible parties, deadlines, and reporting obligations — rather than as accomplished facts or abstract pledges.

Recommendation 2: Submit AI systems to independent third-party auditing before deploying them in high-stakes contexts. The Amazon Rekognition case is the clearest illustration of the communicative consequences of relying on self-reported accuracy claims. hen the ACLU and MIT researchers published empirical evidence of racial and gender bias in Rekognition's outputs, Amazon's rebuttal — that the researchers had used an incorrect confidence threshold — was structurally incapable of resolving the evidentiary dispute, because the company's accuracy assertions rested exclusively on internal testing that had not been verified by a neutral third party. The absence of independent validation meant that Amazon could neither confirm nor credibly dispute the external findings, reducing its crisis communication to a contestation of methodology rather than a substantive engagement with the bias question. The communicative and reputational consequences were directly visible in the data: Amazon received the lowest accountability scores of the three cases ($M = 2.13$), and the mean difference between Amazon and the other two conditions exceeded one full scale point on a five-point instrument.

Companies that deploy AI systems in high-stakes contexts — including law enforcement identification, hiring and recruitment, healthcare decision-support, and public surveillance — should commission independent bias audits from accredited third-party evaluators and publish the results before deployment, regardless of whether external criticism has been published. This practice would do three things simultaneously: provide evidentiary grounding for crisis communications if bias findings emerge; reduce the adversarial dynamic between companies and academic or civil society researchers; and signal to regulators and the public that the company has accepted external accountability as a condition of deployment, rather than as a concession extracted by crisis. In the absence of such audits, any crisis communication claiming that an AI system is accurate, fair, or unbiased will be received as self-interested assertion rather than evidence.

Recommendation 3: Establish governance mechanisms for AI ethics review before deployment decisions are made. The Google Project Maven case demonstrates with particular clarity that the absence of a formal internal review process for ethically sensitive AI deployments creates the structural conditions for avoidable crises. Google's AI Principles — which explicitly prohibited weapons-related AI applications — were published in June 2018, after the Maven contract had already been signed, partially executed, and publicly exposed. The

principles were therefore a crisis response document rather than a governance instrument: they codified restrictions that, had they existed earlier, might have prevented the controversy from arising in the first place. The sequence of events — deployment first, ethics governance second — is precisely the pattern that ethics-washing critics identify as symptomatic of governance designed for reputational management rather than genuine oversight.

Big Tech firms should establish independent AI ethics review boards with real authority — specifically, the power to require modifications to proposed AI applications or reject deployment proposals outright — before controversial deployments are made, not after. These boards should be structurally distinct from marketing and public relations functions, should include external members with relevant technical and domain expertise who are not subject to the company's commercial priorities, and should publish their governance scope, procedures, and decision rationale in sufficient detail to enable external accountability. The critical design requirement is that the board's recommendations must be binding or, where they are overridden, the override decision must itself be subject to governance review and disclosure. Ethics boards whose recommendations are non-binding and whose deliberations are confidential function as legitimacy theatre rather than governance.

Recommendation 4: In governance crises, communicate the specific nature of the failure and the structural changes designed to address it. OpenAI's board statement — that Altman had not been consistently candid in a manner that hindered the board's oversight responsibilities — generated a narrative vacuum that was interpreted in contradictory ways by virtually every stakeholder group it reached. Employees read it as evidence of a personality conflict that posed no structural threat to the company's safety mission; investors read it as a governance malfunction that threatened the company's commercial trajectory; civil society observers read it as confirmation that the non-profit safety mission had been captured by commercial interests. The vagueness of the initial statement was almost certainly deliberate — specific allegations carry legal and relational risks — but the communicative cost of that vagueness was a crisis that lasted five days and resulted in 702 employees threatening resignation, a dynamic that produced the very market instability the communication sought to prevent.

Where legally permissible and practically feasible, companies facing governance crises should communicate the substantive nature of the failure with as much specificity as possible, and should pair that communication with concrete, verifiable commitments about the institutional changes being made in response. Concreteness serves two functions simultaneously: it reduces the space for competing interpretations that amplify uncertainty, and it creates a public commitment record against which the company can subsequently be held

accountable. If specific details cannot be disclosed — due to pending litigation, contractual confidentiality, or regulatory process — companies should explain the reason for that constraint directly, rather than leaving audiences to infer it, and should commit to a specific future date or trigger condition at which further disclosure will occur.

Recommendation 5: Engage with affected communities, not just with regulators and investors. Across all three cases, the communities most directly harmed or threatened by the AI systems under controversy — military personnel and civilian populations affected by drone targeting assistance (Google), independent authors and visual artists whose intellectual property was used without consent or compensation (OpenAI), and communities of colour subjected to inaccurate and racially biased facial recognition in law enforcement contexts (Amazon) — received the least direct, substantive, and stakeholder-specific communication from the companies involved. All three firms concentrated their crisis communications on institutional audiences: regulators, investors, major media partners, and their own employees. This communicative pattern reflects a broader tendency in organisational crisis communication to prioritise the audiences whose power most directly threatens the company's commercial and regulatory position, at the expense of those whose dignity and safety are most directly implicated by the AI system's behaviour.

This tendency is strategically short-sighted as well as ethically problematic. The qualitative analysis showed that the absence of direct community engagement became a recurring source of secondary criticism and sustained legitimacy pressure across all three cases, particularly in the Amazon Rekognition controversy, where civil rights organisations continued to generate damaging publicity for years after the initial crisis precisely because the company had never substantively engaged with the communities most affected by its facial recognition sales to law enforcement. Crisis communication that addresses affected communities directly — acknowledging the specific nature of the harm they experienced, explaining what the company is doing to remedy it, and creating accessible channels for further engagement — is not merely an ethical obligation; it is a legitimacy-repair mechanism that the institutional communication approach structurally cannot replicate, because legitimacy deficits created by harm to specific communities can only be closed by communication directed at those communities.

CONCLUSIONS

This study set out to analyse the crisis communication of Big Tech firms during AI-related ethical scandals and to assess which response elements strengthen stakeholder perceptions of trustworthiness, accountability, and legitimacy. The following conclusions are structured in accordance with the four objectives set in the Introduction.

1. Theoretical conclusions. The theoretical analysis confirms that existing crisis communication frameworks — including Situational Crisis Communication Theory (SCCT), the Social-Mediated Crisis Communication (SMCC) model, and legitimacy-oriented approaches — require substantive extension to account for the structural characteristics of AI ethical scandals. These characteristics include contested and diffuse responsibility attribution arising from the opacity of algorithmic systems; the "many hands" problem, which allows organisations to deflect accountability across technical layers and external actors; platform-mediated amplification, which accelerates crisis escalation and enables competing interpretations to circulate beyond the organisation's control; and an elevated risk of ethics-washing, whereby explicit ethical communication can simultaneously drive positive evaluation and heighten stakeholder scepticism about its genuineness. The AI-specific response typology identified by Wen and Holweg (2024) — deflection, improvement, validation, and pre-emption — provides a useful analytical framework for classifying organisational behaviour, but the present study demonstrates that strategy position alone does not determine perceived credibility. Credibility in AI crisis communication depends less on abstract ethical language and more on answerability, corrective action, governance reform, and stakeholder-oriented communication. The concept of the "double crisis" — in which Big Tech firms must simultaneously defend their capability reputation and their character reputation — remains the central theoretical contribution of this study to the existing literature.

2. Methodological conclusions. The mixed-method design combining qualitative comparative case analysis with a quantitative vignette-based questionnaire proved well-suited to the research problem. Three cases were analysed — Google Project Maven, the OpenAI governance and copyright controversy, and Amazon Rekognition — using directed content analysis across five coding categories. The quantitative strand covered N = 389 respondents (Google: n = 128; OpenAI: n = 130; Amazon: n = 131), satisfying both the power-based minimum (n = 246 for 80% power, $\alpha = 0.05$, medium effect) and the precision-based threshold (95% confidence level, ±5% margin of error, minimum n = 385). Cronbach's Alpha values ranged from 0.668 to 0.799, confirming acceptable internal consistency. Triangulation of both strands enabled a two-sided account of crisis communication: how it is constructed

organisationally and how it is evaluated by stakeholders.

3. Empirical conclusions. The three firms adopted markedly different strategies with different credibility outcomes. Google's pre-emptive response — contract withdrawal paired with published AI Principles — was the most structurally substantive, though partially undermined by a subsequent infrastructure transfer to a third-party contractor. OpenAI's improvement-and-validation response prioritised commercial continuity and carried moderate-to-high ethics-washing risk. Amazon's deflection strategy was the weakest: it denied documented bias, issued non-binding guidelines, and adopted a moratorium only under sustained social pressure. Quantitatively, all three case conditions differed significantly on all four constructs (all $p < .001$). OpenAI received the highest scores on trustworthiness ($M = 3.79$), accountability ($M = 3.80$), and sincerity ($M = 3.65$); Amazon the lowest. All three hypotheses were supported. The cross-cutting finding — that ethics-washing perception correlated positively with all favourable constructs ($r = .373–.419$) — shows that stakeholders can simultaneously perceive a response as credible and as strategically self-protective.

4. Practical conclusions and recommendations. The integrated analysis produces five evidence-based recommendations. First, reform signals should be aligned with operational actions before they are communicated publicly, so that announced commitments are not immediately contradicted by observable corporate behaviour. Second, AI systems deployed in high-stakes contexts should be independently audited by accredited third parties before deployment, providing evidentiary grounding for any subsequent crisis communication. Third, pre-deployment AI ethics governance mechanisms should be established with binding authority rather than advisory remits, ensuring that ethical constraints are built into decision-making rather than appended as post-crisis responses. Fourth, governance crises should be communicated with as much specificity as legally permissible, paired with verifiable institutional commitments, to prevent narrative vacuums from amplifying uncertainty. Fifth, crisis communication should directly and substantively address the communities most affected by the AI system under controversy, rather than concentrating exclusively on regulatory and investor audiences.

Elements of novelty. This study makes three contributions to the existing literature. First, it proposes the concept of the accommodativeness-credibility gap — the finding that the formal accommodativeness of a crisis response strategy does not reliably predict stakeholder-perceived credibility, because visibility and communicative elaborateness mediate the relationship between strategy type and evaluation. Second, it demonstrates empirically that ethics-washing perception is not the inverse of positive crisis response evaluation but co-varies with it, indicating that the same communicative features that build credibility also heighten

scepticism about genuineness. Third, it provides an integrated empirical design that connects actual corporate crisis communication texts with audience-level evaluations of those texts in a single study — an approach that has been called for in the literature but remains underrepresented in empirical work on AI-related ethical scandals.

Possibilities and restrictions of applying the results. The practical recommendations derived from this study are directly applicable to communication practitioners, AI ethics teams, and governance officers in large technology organisations facing or preparing for AI-related controversies. The empirical framework — distinguishing substantive from symbolic reform signals and measuring their reception across four stakeholder evaluation constructs — can be adapted for internal crisis communication audits and pre-crisis communication planning. The restrictions on application follow from the study's methodological limitations: the convenience and snowball sample is not statistically representative of any specific national or professional population, and the vignette design, while enabling controlled comparison, simplifies the complexity of real-world reputation formation. The findings should therefore be treated as empirically grounded directional evidence rather than universally generalisable prescriptions, and their application in specific organisational contexts should be accompanied by stakeholder-specific research.

Suggestions for further research. Several directions for future research follow from the limitations and findings of this study. First, experimental designs in which individual crisis communication variables — stakeholder orientation, accountability signal specificity, channel choice — are manipulated independently would allow more precise causal attribution than the vignette design used here permits. Second, longitudinal research tracking stakeholder trust over the full arc of an AI crisis — from initial disclosure through governance response to post-crisis reputation — would address the cross-sectional limitation of the present study and capture how perceptions shift as new information emerges. Third, cross-cultural comparative research is needed to assess whether the patterns identified here — particularly the positive correlation between ethics-washing perception and favourable constructs — generalise across different regulatory, cultural, and media environments. Fourth, future research could extend the case selection beyond three firms and three scandal types to test whether the accommodativeness-credibility gap is a stable feature of AI crisis communication or specific to the cases analysed here. Finally, research examining the internal governance conditions that produce proactive versus reactive disclosure — including the role of ethics board authority, internal reporting structures, and regulatory pressure — would complement the external communication focus of the present study.

REFERENCES

Biddle, J. B., Nelson, J. P., & Olugbade, O. E. (2024). How Can We Know if You are Serious? Ethics Washing, Symbolic Ethics Offices, and the Responsible Design of AI Systems. *Canadian Journal of Philosophy*, *54*(4), 329–345. https://doi.org/10.1017/can.2025.9

Che, S., Zhou, Y., Zhang, S., Nan, D., & Kim, J. H. (2023). Impact of ByteDance crisis communication strategies on different social media users. *Humanities and Social Sciences Communications*, *10*(1), 729. https://doi.org/10.1057/s41599-023-02170-3

Cheng, Y., Li, J., Lee, J., & Liu, Y. (2025). How is artificial intelligence shaping crisis communication? A systematic review and future research agenda. *Journal of Risk Research*, 1–21. https://doi.org/10.1080/13669877.2025.2579317

Cheng, Y., Wang, Y., & Kong, Y. (2022). The state of social-mediated crisis communication research through the lens of global scholars: An updated assessment. *Public Relations Review*, *48*(2), 102172. https://doi.org/10.1016/j.pubrev.2022.102172

Cheong, B. C. (2024). Transparency and accountability in AI systems: Safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, *6*, 1421273. https://doi.org/10.3389/fhumd.2024.1421273

Coombs, W. T. (2023). *Ongoing crisis communication: Planning, managing, and responding* (Sixth edition). SAGE.

Coombs, W. T., & Tachkova, E. R. (2022). Elaborating the concept of threat in contingency theory: An integration with moral outrage and situational crisis communication theory. *Public Relations Review*, *48*(4), 102234. https://doi.org/10.1016/j.pubrev.2022.102234

De Almeida, P. G. R., & Dos Santos Júnior, C. D. (2025). Artificial intelligence governance: Understanding how public organizations implement it. *Government Information Quarterly*, *42*(1), 102003. https://doi.org/10.1016/j.giq.2024.102003

Eisele, O., Tolochko, P., & Boomgaarden, H. G. (2022). How do executives communicate about

crises? A framework for comparative analysis. *European Journal of Political Research*, *61*(4), 952–972. https://doi.org/10.1111/1475-6765.12504

Floridi, L., Holweg, M., Taddeo, M., Amaya Silva, J., Mökander, J., & Wen, Y. (2022). capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4064091

Funmilola Olatundun Olatoye, Kehinde Feranmi Awonuga, Noluthando Zamanjomane Mhlongo, Chidera Victoria Ibeh, Oluwafunmi Adijat Elufioye, & Ndubuisi Leonard Ndubuisi. (2024). AI and ethics in business: A comprehensive review of responsible AI practices and corporate responsibility. *International Journal of Science and Research Archive*, *11*(1), 1433–1443. https://doi.org/10.30574/ijsra.2024.11.1.0235

Hornsey, M. J., Chapman, C. M., La Macchia, S., & Loakes, J. (2024). Corporate apologies are effective because reform signals are weighted more heavily than culpability signals. *Journal of Business Research*, *177*, 114620. https://doi.org/10.1016/j.jbusres.2024.114620

Huang, K., & Krafft, P. M. (2024). Performing Platform Governance: Facebook and the Stage Management of Data Relations. *Science and Engineering Ethics*, *30*(2), 13. https://doi.org/10.1007/s11948-024-00473-5

Huang, M., & Ki, E.-J. (2023). Protecting organizational reputation during a para-crisis: The effectiveness of conversational human voice on social media and the roles of construal level, social presence, and organizational listening. *Public Relations Review*, *49*(5), 102389. https://doi.org/10.1016/j.pubrev.2023.102389

Jang, E., & Plaisance, P. L. (2025). Textual and Comparative Analyses on AI Policies: How Big Tech and Their Global Watchdogs Frame Virtues and Responsibility*. *Journal of Media Ethics*, *40*(4), 187–204. https://doi.org/10.1080/23736992.2025.2580662

Jong, W. (2025). Beyond the snapshot: Rethinking crisis communication theories in dynamic

crisis situations. *Public Relations Review*, *51*(3), 102586. https://doi.org/10.1016/j.pubrev.2025.102586

Keymolen, E. (2024). Trustworthy tech companies: Talking the talk or walking the walk? *AI and Ethics*, *4*(2), 169–177. https://doi.org/10.1007/s43681-022-00254-5

Lim, J. S., Lee, C., Shin, D., Kim, J., & Zhang, J. (2025). Perceived Stakeholder Engagement in Corporate Data Responsibility (CDR) Communication and Its Relationship with Trust in Generative AI Systems: The Mediating Role of Algorithmic and Institutional Responsibility. *Journal of Public Relations Research*, *37*(5), 447–469. https://doi.org/10.1080/1062726X.2025.2501552

Mehta, S., & Erickson, K. (2022). Can online political targeting be rendered transparent? Prospects for campaign oversight using the Facebook Ad Library. *Internet Policy Review*, *11*(1). https://doi.org/10.14763/2022.1.1648

Novelli, C., Taddeo, M., & Floridi, L. (2024). Accountability in artificial intelligence: What it is and how it works. *AI & SOCIETY*, *39*(4), 1871–1882. https://doi.org/10.1007/s00146-023-01635-y

Opgenhaffen, M. (2023). Combatting disinformation with crisis communication : An analysis of Meta's newsroom stories. *Communications*, *48*(3), 352–369. https://doi.org/10.1515/commun-2022-0101

Page, T. G., Zhou, A., & Capizzo, L. W. (2023). Finding the path beyond reputation repair: A structural topic modeling analysis of the crisis communication paradigm in public relations. *Public Relations Review*, *49*(4), 102349. https://doi.org/10.1016/j.pubrev.2023.102349

Qu, J. G., Yi, J., Zhang, W. J., & Yang, C. Y. (2023). Silence is golden? Mitigating different types of online firestorms of Fortune 100 corporations on Twitter. *Public Relations Review*, *49*(5), 102391. https://doi.org/10.1016/j.pubrev.2023.102391

Ray, E. C., Merle, P. F., & Lane, K. (2025). Generating Credibility in Crisis: Will an AI-

Scripted Response Be Accepted? *International Journal of Strategic Communication*, *19*(2), 158–175. https://doi.org/10.1080/1553118X.2024.2435494

Reid, A., & Ringel, E. (2025). Digital intermediaries and transparency reports as strategic communications. *The Information Society*, *41*(2), 91–109. https://doi.org/10.1080/01972243.2025.2453529

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. *Organizational Behavior and Human Decision Processes*, *188*, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Trautman, L. J. (2024). Sam Altman, OpenAI, and the Importance of Corporate Governance. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4679613

Van Maanen, G. (2022). AI Ethics, Ethics Washing, and the Need to Politicize Data Ethics. *Digital Society*, *1*(2), 9. https://doi.org/10.1007/s44206-022-00013-3

Venturini, T. (2010). Diving in magma: How to explore controversies with actor-network theory. *Public Understanding of Science*, *19*(3), 258–273. https://doi.org/10.1177/0963662509102694

Wen, Y., & Holweg, M. (2024). A phenomenological perspective on AI ethical failures: The case of facial recognition technology. *AI & SOCIETY*, *39*(4), 1929–1946. https://doi.org/10.1007/s00146-023-01648-7

Yoganathan, V., Osburg, V.-S., & Janakiraman, N. (2025). Lending Legitimacy to Corporate Digital Responsibility: Trust in Firm Versus Government Regulation of Artificial Intelligence Services. *Journal of Service Research*, 10946705251345097. https://doi.org/10.1177/10946705251345097

APPENDICES

Appendix 1

Glossary of terms

Accountability — the obligation of an organisation to answer for its actions, explain its decisions, and accept consequences for AI-related harms. In AI contexts, accountability extends beyond legal compliance to encompass transparent governance structures, verifiable corrective action, and enforceable oversight mechanisms (Novelli, Taddeo, & Floridi, 2024).

AI ethical scandal — a publicly documented controversy in which an AI application is perceived to violate social norms or values, escalating beyond technical failure into a legitimacy conflict involving media, civil society, and regulatory actors. AI ethical scandals are characteristically complex because responsibility is often difficult to attribute clearly due to the opacity of algorithmic decision-making (Wen & Holweg, 2024).

Algorithmic opacity — the characteristic difficulty of explaining the internal decision-making logic of AI systems, particularly deep learning models, to non-technical audiences. Algorithmic opacity complicates crisis communication because organisations cannot easily provide the clear, verifiable explanations that stakeholders expect when AI systems cause harm (Wen & Holweg, 2024).

Answerability — a structural property of genuine accountability requiring that an organisation can be questioned, must provide substantive explanations, and is subject to enforceable consequences. Answerability distinguishes substantive crisis responses from purely rhetorical reassurances (Novelli, Taddeo, & Floridi, 2024).

Big Tech — large, globally operating technology companies with significant market power, extensive platform infrastructures, and substantial AI development capacity; in this study refers specifically to Google, Amazon, and OpenAI, with broader reference to companies such as Meta, Apple, and Microsoft (Huang & Krafft, 2024).

Character reputation — a stakeholder judgement concerning whether an organisation's intentions and values are appropriate and trustworthy, as distinct from its technical capabilities. In AI ethical scandals, character reputation is threatened when crisis responses appear self-protective rather than accountability-oriented (Wen & Holweg, 2024).

Crisis communication — the collection, processing, and dissemination of information required to address a crisis situation, with the goals of reducing uncertainty, protecting organisational reputation, and restoring stakeholder trust. In this study, crisis communication is analysed as an information management intervention, a reputational repair mechanism, and a legitimacy management process (Coombs, 2023).

Deflection — a crisis response strategy in which an organisation attributes failures to external parties, customer misuse, or methodological errors in external research rather than acknowledging inherent flaws in its systems. Deflection is the least accommodative strategy on the crisis response spectrum (Wen & Holweg, 2024).

Ethics-washing — the practice of publicly signalling a commitment to ethics through statements, principles, or governance structures without acting in ways sufficiently aligned with that signalling, and in the absence of enforceable oversight or reporting obligations. Ethics-washing transforms moral language into a reputational management tool rather than a genuine accountability mechanism (Van Maanen, 2022; Biddle, Nelson, & Olugbade, 2024).

Legitimacy repair — the communicative and structural processes through which an organisation attempts to restore alignment between its behaviour and the norms, values, and expectations of its stakeholders following a reputational crisis. Legitimacy repair strategies range from moral positioning and reform narratives to substantive governance restructuring (Hornsey, Chapman, La Macchia, & Loakes, 2024).

Pre-emption — a crisis response strategy in which an organisation takes proactive measures — such as voluntarily withdrawing a controversial product or publishing binding ethical principles — before a crisis reaches its peak and stakeholder blame fully consolidates. Pre-emption is the most accommodative strategy on the crisis response spectrum (Wen & Holweg, 2024).

Sincerity (perceived) — a stakeholder evaluation construct measuring the degree to which a corporate crisis response is judged to be genuine and honest rather than strategically calculated or performative (Hornsey, Chapman, La Macchia, & Loakes, 2024).

Situational Crisis Communication Theory (SCCT) — a theoretical framework prescribing that crisis response strategies should be matched to the degree of responsibility stakeholders attribute to the organisation, classifying crises along a continuum from victim through accidental to preventable crises and recommending progressively more accommodative responses as attributed responsibility increases (Coombs, 2023).

Social-Mediated Crisis Communication (SMCC) — a model extending SCCT to account for social media dynamics, demonstrating that message reception depends on the channel through which it travels and the audience encountering it. The SMCC model is relevant to Big Tech crises because the same corporate communication can be evaluated very differently across owned and independent platforms (Cheng, Wang, & Kong, 2022).

Appendix 2

Declaration on the usage of generative artificial intelligence (GAI) and GAI-based tools (GAI tools)

24.04.2026

Inna Myroniuk
CRISIS COMMUNICATION OF BIG TECH FIRMS DURING AI-RELATED ETHICAL SCANDALS

☐ **I confirm that NO GAI tools were used in the preparation of this work.**

☑ **I confirm that the GAI tools listed and described below were used in the preparation of this work :**

GAI tool (name and version, e.g. ChatGPT version 3.5): __ ChatGPT, version GPT-3.5_____

Company that developed the GAI tool (e.g., OpenAI):___ OpenAI_____

GAI tool web address (e.g., https://openai.com/chatgpt/):__ https://openai.com/chatgpt/_____

Date of content/output creation (e.g., xx/xx/20xx): ___ 03/2026 – 04/2026_____

Purpose of using the GAI tool (e.g., data collection, editing, translation, or other):___Language editing and grammar correction of draft sections written by the author; rewording of individual sentences for clarity and academic register; translation support for terminology verification between Ukrainian and English.__

Specific location (section/page) where the GAI tool was used:___Introduction; Chapter 1 (literature review sections); Chapter 2 (methodology sections); selected paragraphs throughout Chapter 3 for language editing purposes

Continuation of Appendix 2

Declaration on the usage of generative artificial intelligence (GAI) and GAI-based tools (GAI tools)

24.04.2026

Inna Myroniuk
CRISIS COMMUNICATION OF BIG TECH FIRMS DURING AI-RELATED ETHICAL SCANDALS

☐ **I confirm that NO GAI tools were used in the preparation of this work.**

☑ **I confirm that the GAI tools listed and described below were used in the preparation of this work :**

GAI tool (name and version, e.g. ChatGPT version 3.5): Claude, version Claude Sonnet 4.6

Company that developed the GAI tool (e.g., OpenAI): Anthropic

GAI tool web address (e.g., https://openai.com/chatgpt/): https://claude.ai

Date of content/output creation (e.g., xx/xx/20xx): 04/2026 – 05/2026

Purpose of using the GAI tool (e.g., data collection, editing, translation, or other): Language editing, academic register refinement, and structural feedback on draft sections written by the author; assistance with rewriting and improving clarity of passages in Chapter 2 and Chapter 3; support with formatting consistency and wording of the quantitative results narrative

Specific location (section/page) where the GAI tool was used: Chapter 2 (Sections 2.4, 2.5, 2.6); Chapter 3 (Sections 3.2, 3.3); Conclusions; Glossary of Terms (Annex 1)

Appendix 3

SURVEY QUESTIONNAIRE

Introduction text (shown to participants): This survey is part of a master's thesis research project aiming to identify how people react to crisis communication of big tech firms during AI-related ethical scandals. It takes approximately 8–10 minutes. You will read a short scenario about technology company involved in a controversy, then answer questions about how you perceive the company's response. There are no right or wrong answers — your personal opinion is what matters. All responses are anonymous.

SECTION 1 — BACKGROUND (screening + demographics)

1. Your age group:
   - 18–24 / 25–34 / 35–44 / 45–54 / 55+
2. Your gender
   - Male/Female/Other/Don't want to disclose
3. Your country of residence: *(open field)*
4. Your highest level of education:
   - Secondary / Bachelor's / Master's / PhD / Other
5. How often do you use AI-powered tools or services? (e.g., ChatGPT, Google Assistant, Netflix recommendations, Spotify, face filters)
   - Daily
   - Several times a week
   - Occasionally
   - Rarely or never
6. Are you aware that many everyday apps use artificial intelligence?
   - Yes, I know quite a lot about it
   - I have a general idea
   - Not really
   - No
7. Which of the following AI tools have you used? (select all that apply)

   ☐ Chatbots (e.g., ChatGPT, Gemini, Copilot)

   ☐ Voice assistants (Siri, Alexa, Google Assistant)

   ☐ Recommendation systems (Netflix, Spotify, YouTube)

   ☐ AI image or video tools (Midjourney, DALL-E, filters)

☐ AI writing or productivity tools

☐ None of the above

8. Have you ever seen news about an AI company being criticized for causing harm? (e.g., biased algorithms, privacy violations)

- Yes, I remember specific cases
- Yes, I have heard about this generally
- No
- Not sure

SECTION 2 — YOUR GENERAL VIEWS ON BIG TECHNOLOGY COMPANIES

9. Please indicate how much you agree with the following statements on the scale from 1 to 5, where 1 – strongly disagree, and 5 – strongly agree.

	1	2	3	4	5
I generally trust major big technology companies.	○	○	○	○	○
I am skeptical of big technology companies' public claims about AI ethics.	○	○	○	○	○
Big technology companies should be externally monitored when they use AI in high-impact decisions.	○	○	○	○	○
Big technology companies should clearly explain what went wrong when their AI systems cause harm.	○	○	○	○	○
I trust a big technology company more when it admits specifically what went wrong.	○	○	○	○	○
I trust a big technology company more when it takes concrete steps to fix the problem.	○	○	○	○	○
A response of big technology company with only general statements about values feels insufficient to me.	○	○	○	○	○
I can usually tell the difference between a genuine apology of a big technology company and a scripted one.	○	○	○	○	○
Most big technology companies' public commitments to ethics are mainly for show.	○	○	○	○	○
Big technology companies should be legally required to explain how their AI makes decisions.	○	○	○	○	○
Independent regulators should check whether big tech companies keep their promises.	○	○	○	○	○

10. *What would you personally do in ?* Please indicate how much you agree with the following statements on the scale from 1 to 5, where 1 – strongly disagree, and 5 – strongly agree.

	1	2	3	4	5
If a big technology company's AI caused discrimination, I would stop using its services.	○	○	○	○	○
I would share news about an big technology company's misconduct on social media.	○	○	○	○	○
I would support stricter regulation of big technology companies after learning about harm they caused.	○	○	○	○	○

SECTION 3 — VIGNETTE SCENARIO

GOOGLE

In 2018, it was publicly revealed that company X (Google) had partnered with the United States Department of Defense on a military programme that used AI to analyse drone surveillance footage. The revelation triggered a public petition signed by over 3,000 company (Google) employees and sparked widespread media debate about the ethical use of AI in military contexts. Employees argued the project contradicted companie's (Google's) stated values.

Now you will see Google's official response. After reading it, you will be asked for your opinion.

Google's official response:

Companie's CEO (Google's then-CEO Sundar Pichai) published the company's AI Principles, which explicitly stated that they (Google) would not develop AI for use in weapons or surveillance that violates international norms. The company subsequently did not renew its project contract when it expired in 2019 and announced it would not pursue similar military AI contracts in the future.

OPENAI

Background: In late 2023 and early 2024, Company Y (OpenAI) faced significant public controversy on two fronts. First, a group of senior employees and board members raised concerns about the pace of AI development and insufficient safety oversight, leading to a high-profile internal governance crisis. Second, multiple artists and media organisations publicly accused Company Y of training its AI models on copyrighted content without consent or compensation, raising ethical questions about data practices and intellectual property.

Now you will see OpenAI's official response. After reading it, you will be asked for your opinion.

Official response:

Company Y published a statement reaffirming its commitment to safety and responsible AI development. It announced the formation of a new Safety and Security Committee
and committed to publishing regular safety updates. On the copyright issue, Company Y
acknowledged ongoing legal proceedings and stated it was working with content creators on licensing arrangements, while maintaining that its training practices were consistent with fair use principles.

AMAZON

Background: Company Z's (Amazon's) facial recognition tool, Rekognition, was publicly criticised from 2018 onwards after studies — including one by MIT researchers — demonstrated that the system was significantly less accurate in identifying the gender of darker-skinned women compared to lighter-skinned men. The American Civil Liberties Union (ACLU) also found that Rekognition incorrectly matched 28 members of the US Congress with criminal mugshots. Civil rights organisations called on Amazon to stop selling the technology to law enforcement agencies.

Now you will see Amazon's official response. After reading it, you will be asked for your opinion.

Official response:

Company Z initially deflected criticism by arguing that the technology was not inherently problematic and that any issues resulted from incorrect use by customers. The company published usage guidelines for law enforcement and called for federal legislation to govern facial recognition. In June 2020, following the Black Lives Matter movement, Company Z announced a one-year moratorium on police use of Rekognition, which was later extended indefinitely

Your reaction to the response:

Please indicate how much you agree with the following statements on the scale from 1 to 5, where 1 – strongly disagree, and 5 – strongly agree.

	1	2	3	4	5
Company X is being honest about what happened.	○	○	○	○	○
After reading this, I feel more confident about Company X.	○	○	○	○	○
I believe Company X will actually follow through on its promises.	○	○	○	○	○

	1	2	3	4	5
Company X clearly takes responsibility for what went wrong.	○	○	○	○	○
Company X explains what will be done to prevent this from happening again.	○	○	○	○	○
The response includes concrete and verifiable commitments, not just general promises.	○	○	○	○	○

	1	2	3	4	5
This response feels genuine rather than scripted.	○	○	○	○	○
I believe Company X truly wants to fix the problem.	○	○	○	○	○

	1	2	3	4	5
This response feels like a way to avoid real accountability.	○	○	○	○	○
Company X seems more interested in protecting its image than fixing the problem.	○	○	○	○	○
This statement feels like it was written mainly for public relations purposes.	○	○	○	○	○

1. Overall, how would you rate Company X response?
   - Very poor
   - Poor
   - Neither good nor poor
   - Good
   - Very good
2. After reading Company X response, how likely are you to trust this company?
   - Much less likely to trust them
   - Slightly less likely
   - No change
   - Slightly more likely
   - Much more likely to trust the

Appendix 4

Scale Reliability Statistics (Cronbach's α)

Table A4.1.

Summary of internal consistency for all four constructs.

Scale	Cronbach's α	No. of items
Trustworthiness	0.799	3
Accountability	0.739	3
Sincerity	0.668	2
Ethics-washing perception	0.703	3

Note. All α values meet the acceptable threshold (α ≥ 0.60). Source: compiled by the author.

Table A4.2.

Item-level reliability statistics — Trustworthiness (α = 0.799, items V–X).

Item	Item-rest correlation	Cronbach's α if item dropped
V	0.734	0.622
W	0.563	0.805
X	0.642	0.726

Note. Source: compiled by the author.

Table A4.3.

Item-level reliability statistics — Accountability (α = 0.739, items Y–AA).

Item	Item-rest correlation	Cronbach's α if item dropped
Y	0.559	0.660
Z	0.606	0.603
AA	0.528	0.696

Note. Source: compiled by the author.

Table A4.4.

Item-level reliability statistics — Sincerity (α = 0.668, items AB–AC).

Item	Item-rest correlation	Cronbach's α if item dropped
AB	0.502	0.487
AC	0.502	0.517

Note. Source: compiled by the author.

Table A4.5.

Item-level reliability statistics — Ethics-washing perception (α = 0.703, items AD–AF).

Item	Item-rest correlation	Cronbach's α if item dropped
AD	0.514	0.620
AE	0.568	0.550
AF	0.480	0.660

Note. Source: compiled by the author.

Appendix 5

Pearson Correlation Matrix

Table A5.1.

Pearson correlations between construct mean scores (N = 389).

		Trust_Mean	*Acc_Mean*	*Sinc_Mean*	*Ethics_Mean*
Trust_Mean	*Pearson's r*	—			
	95% CI Upper	—			
	95% CI Lower	—			
Acc_Mean	*Pearson's r*	*0.771***	—		
	95% CI Upper	*0.809*	—		
	95% CI Lower	*0.728*	—		
Sinc_Mean	*Pearson's r*	*0.685***	*0.670***	—	
	95% CI Upper	*0.735*	*0.722*	—	
	95% CI Lower	*0.629*	*0.612*	—	
Ethics_Mean	*Pearson's r*	*0.373***	*0.419***	*0.313***	—
	95% CI Upper	*0.455*	*0.497*	*0.400*	—
	95% CI Lower	*0.284*	*0.333*	*0.220*	—
*Note. * p < .05, ** p < .01, *** p < .001*					

Note. Source: compiled by the author.

Appendix 6

Descriptive Statistics and One-Way ANOVA Results

Table A6.1.

Descriptive statistics by case condition (1 = Google, 2 = OpenAI, 3 = Amazon).

Statistic	Case	Trust_Mean	Acc_Mean	Ethics_Mean	Sinc_Mean
N	1	128	128	128	128
	2	130	130	130	130
	3	131	131	131	131
Mean	1	3.24	3.17	3.18	3.10
	2	3.79	3.80	3.74	3.65
	3	2.02	2.13	2.84	2.31
Median	1	3.33	3.00	3.33	3.00
	2	4.00	4.00	4.00	4.00
	3	2.00	2.00	2.67	2.00
SD	1	0.654	0.695	0.663	0.811
	2	0.752	0.735	0.797	0.935
	3	0.708	0.643	1.010	0.858
Min	1	1.33	1.33	1.00	1.00
	2	1.33	1.33	1.00	1.00
	3	1.00	1.00	1.00	1.00
Max	1	5.00	5.00	5.00	5.00
	2	5.00	5.00	5.00	5.00
	3	4.00	4.00	5.00	5.00

Note. Case 1 = Google (Project Maven); Case 2 = OpenAI (Governance & Copyright); Case 3 = Amazon (Rekognition). Source: compiled by the author.

Table A6.2.

One-Way ANOVA summary.

Dependent variable	Test	F	df1	df2	p
Trust_Mean	Fisher's	215	2	386	< .001
Acc_Mean	Fisher's	195	2	386	< .001
Sinc_Mean	Fisher's	78.8	2	386	< .001
Ethics_Mean	Welch's	35.7	2	252	< .001
Ethics_Mean	Fisher's	38.7	2	386	< .001

Note. Ethics_Mean violated homogeneity of variances (Levene's $F = 18.6$, $p < .001$); Welch's F reported as primary statistic. Source: compiled by the author.

Table A6.3.

Levene's test for homogeneity of variances.

Variable	F	df1	df2	p
Trust_Mean	0.587	2	386	.556
Acc_Mean	0.474	2	386	.623
Sinc_Mean	0.980	2	386	.376
Ethics_Mean	18.6	2	386	< .001

Note. Source: compiled by the author.

Table A6.4.

Tukey post-hoc test — Trust_Mean.

		Case 1	**Case 2**	**Case 3**
Case 1	Mean difference	—	-0.555	1.21
	p-value	—	< .001	< .001
Case 2	Mean difference		—	1.77
	p-value		—	< .001
Case 3	Mean difference			—
	p-value			—

Note. Source: compiled by the author.

Table A6.5.

Tukey post-hoc test — Acc_Mean.

		Case 1	**Case 2**	**Case 3**
Case 1	Mean difference	—	-0.628	1.04
	p-value	—	< .001	< .001
Case 2	Mean difference		—	1.67
	p-value		—	< .001
Case 3	Mean difference			—
	p-value			—

Note. Source: compiled by the author.

Table A6.6.

Tukey post-hoc test — Sinc_Mean.

		Case 1	**Case 2**	**Case 3**
Case 1	Mean difference	—	-0.552	0.792
	p-value	—	< .001	< .001
Case 2	Mean difference		—	1.345
	p-value		—	< .001
Case 3	Mean difference			—
	p-value			—

Note. Source: compiled by the author.

Table A6.7.

Games-Howell post-hoc test — Ethics_Mean (used due to unequal variances).

		Case 1	**Case 2**	**Case 3**
Case 1	Mean difference	—	-0.564	0.340
	p-value	—	< .001	.004
Case 2	Mean difference		—	0.904
	p-value		—	< .001
Case 3	Mean difference			—
	p-value			—

Note. Source: compiled by the author.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management
Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/lsm**