

Kevin Bauer  
Katja Langenbucher

# Agentic AI, Corporate Communication, and Market Integrity

SAFE Policy Letter No. 113 | May 2026

**Leibniz Institute for Financial Research SAFE**  
Sustainable Architecture for Finance in Europe

[policy\\_center@safe-frankfurt.de](mailto:policy_center@safe-frankfurt.de) | [www.safe-frankfurt.de](http://www.safe-frankfurt.de)

# Agentic AI, Corporate Communication, and Market Integrity\*

Kevin Bauer<sup>†</sup>   Katja Langenbucher<sup>‡</sup>

May 5, 2026

## Summary

Agentic artificial intelligence (AI) is increasingly used in corporate communication with investors, including drafting disclosures, answering queries, and summarizing financial information. While these systems can improve the accessibility and efficiency of public corporate communication, they also create risks for market integrity, such as inaccurate statements, unintended disclosure of sensitive information, and unequal access through personalized responses.

This paper argues that existing disclosure and market-abuse frameworks remain substantively adequate but require clearer operational expectations for AI-driven communication. AI outputs delivered through issuer-controlled channels should be treated as corporate communications attributable to the issuer. A risk-based regulatory approach should therefore require governance oversight, separation of public and confidential data, safeguards against manipulation, auditable records of AI outputs, and human oversight for market-sensitive communications. Properly governed, agentic AI can enhance investor access to public information while preserving the principles of accurate, timely, and equal disclosure.

\*SAFE policy papers represent the authors' personal opinions and do not necessarily reflect the views of the Leibniz Institute for Financial Research SAFE or its staff.

<sup>†</sup>Goethe University Frankfurt (Professor for Game-Theoretic and Causal AI in Business and Economics), Leibniz Institute for Financial Research SAFE, Hessian.AI, Email: [bauer@safe-frankfurt.de](mailto:bauer@safe-frankfurt.de)

<sup>‡</sup>Goethe University Frankfurt (Professor of Civil Law and Financial Market Regulation), Leibniz Institute for Financial Research SAFE, Institute for Monetary and Financial Stability (IMFS), Email: [langenbucher@safe-frankfurt.de](mailto:langenbucher@safe-frankfurt.de)

## Executive Summary

Agentic artificial intelligence (AI) is beginning to play a role in corporate communication with investors. It can materially improve the efficiency and accessibility of preparing and explaining public information. These systems can draft disclosures, answer investor questions, summarize filings, and translate complex information into more accessible language. Used responsibly, they could improve the speed, clarity, and accessibility of public financial information.

However, deploying AI in investor communication also introduces significant risks for market integrity and investor protection. AI systems can generate large volumes of statements quickly, increasing the likelihood of errors, misleading interpretations, or unintended disclosure of sensitive information. Even small inaccuracies in financial communications can affect investor decisions and market prices.

Additional risks arise from personalization and interactive responses. Tailored answers or emphasis could unintentionally provide unequal access to information or enable investors to infer non-public details. Moreover, AI systems interacting in open digital environments may be vulnerable to manipulation attempts designed to override safeguards or extract confidential information.

Existing securities laws already require companies to ensure that market communications are accurate, non-misleading, and publicly accessible. The challenge is therefore operational rather than legal: regulators and companies must translate these obligations into governance and technical controls suitable for AI-driven communication systems.

AI outputs delivered through issuer-controlled channels should be treated as corporate communications attributable to the company, meaning firms remain responsible for their accuracy and compliance with securities law. Regulators should establish clear supervisory expectations for companies deploying investor-facing AI. These expectations should include identifiable governance responsibility, strict separation between public and confidential data sources, safeguards against manipulation and prompt-injection attacks, auditable records of how AI outputs were generated, and meaningful human oversight for communications that could influence markets

A risk-based approach would distinguish between systems that merely explain already-public information and systems capable of generating new market-relevant claims or accessing internal data, with the latter subject to stricter controls. By clarifying these operational expectations, regulators can enable the benefits of AI-assisted communication, such as improved accessibility and investor understanding, while preserving the core principles of accurate, timely, and equal disclosure that underpin market confidence.

## Overview

Regulators are entering a phase of financial communication in which listed companies can deploy agentic AI systems to draft disclosures, respond to investor queries conversationally, monitor market narratives, and execute multi-step workflows that retrieve and synthesize information from both internal and external sources. These capabilities can reduce friction in investor access to public information by improving speed, readability, translation, and consistency. They also amplify long-established securities-law risks by scaling the volume and reach of issuer-attributable statements, increasing the probability of error, and widening pathways through which inside information can be inferred, mishandled, or disclosed. The core policy conclusion is that agentic AI can materially improve the efficiency and accessibility of preparing and explaining public information, but it does not, without stringent governance and system design, reliably support lawful external-facing communication in adversarial environments. The regulatory objective should therefore be to preserve accountability under existing disclosure, market-abuse, and antifraud frameworks while establishing enforceable operational expectations that translate those duties into controls suited to agentic systems.

## Context and Regulatory Landscape

Corporate disclosure regimes are designed to reduce information asymmetry, promote accurate price formation, and protect investors by requiring timely, complete, and non-misleading public information. Across major jurisdictions, issuers must produce audited financial statements under specified standards and provide narrative and event-driven disclosures subject to materiality thresholds and stringent constraints on the handling of inside information. These frameworks rely on attributable processes: identifiable responsible persons, reviewable records, and controlled channels of dissemination. In the European Union, market-abuse rules reinforce this architecture by requiring prompt public disclosure of inside information, constraining misuse, and imposing confidentiality disciplines where disclosure is lawfully delayed. In the United States, while the regulatory structure differs, anti-fraud provisions, exchange obligations, and fair disclosure expectations similarly penalize misleading statements and improper dissemination of market-moving information. In both settings, the law's tolerance for error is low where communications can influence trading behavior and market integrity.

Agentic AI changes the operational conditions under which these duties must be fulfilled. A conversational agent deployed on an investor relations channel can answer continuously, tailor explanation to audience sophistication, translate content, and synthesize information from filings, press releases, earnings materials, and other sources. This alters not only how statements are produced but how information moves, because the agent can generate outputs at high frequency, at scale, and under uncertain user intent. It also creates new vectors for regulatory breach, including inadvertent fabrication, unjustified inference, selective emphasis that enables inference of nonpublic information, and differential access created through personalization. The legal gap is not a lack of substantive prohibitions; the gap is operational. Existing duties are well established, but regulators lack a consistent set of enforceable expectations that specify how issuers must govern and constrain systems that can retrieve, plan, and generate content dynamically in environments that include manipulation attempts, hostile actors, and ambiguous prompts.

## How Agentic AI Alters Disclosure and Market-Integrity Risk

Agentic AI affects compliance risk even when used “only for drafting,” because drafting is an upstream stage in regulated communication and influences what ultimately reaches the market. Language that appears authoritative can migrate into external channels through routine workflows, and the velocity of AI-assisted drafting increases the likelihood that unverified text is reused across disclosures, presentations, social media, and investor messaging. In parallel, market-intelligence uses can create feedback loops in which investor queries, sentiment analytics, and narrative monitoring influence issuer messaging strategy. When such analytics are used to calibrate content, timing, or emphasis for particular audiences, they can undermine equal access to information and increase selective disclosure risk, even where the issuer intends only to improve engagement.

The technical failure modes of language models translate directly into securities-law exposure because plausible text is not equivalent to accurate fact. Hallucinations are not merely occasional mistakes; they can arise at multiple points in an agent’s chain of actions. The system may misinterpret a question, fail to retrieve necessary information, misunderstand retrieved content, or generate inferences that are rhetorically coherent but factually unsupported. In capital markets, the harm is immediate when an output invents a financial figure, misstates a reporting period, implies an unannounced corporate event, or offers an overly confident account of forward-looking impacts. Because agentic systems can plan multi-step actions and propagate outputs across channels, the operational risk is compounded: an initial error can be repeated, reformulated, or amplified before it is detected, increasing the likelihood that it becomes price-relevant and that the issuer’s controls appear deficient.

Risk is not uniform across communication contexts, and regulatory expectations should reflect consistent triggers tied to market impact rather than to marketing labels. Retail-facing deployments can advance investor protection by translating complex filings into plain language, improving comprehension of already-public information, and reducing information overload. Yet these systems also blur the boundary between neutral explanation and promotional framing, and personalization can become misleading when it changes emphasis, certainty, or implied implications beyond what the public record supports. Engagement with sophisticated investors creates a different pressure point, because responsiveness and customization may drift into selective disclosure or preferential access, particularly if the system can draw from internal materials or if its outputs enable inference of nonpublic information through subtle deviations in wording. Adversarial contexts such as responses to market attacks, activist campaigns, or control contests are especially sensitive because incentives for rapid automated response collide with constraints on manipulation and market integrity, and because hostile actors are more likely to test system vulnerabilities.

## Adversarial Resilience as a Disclosure-Control Requirement

Agentic AI introduces a class of compliance risk that is simultaneously a disclosure-control issue and a cybersecurity issue. Prompt injection and related attacks are operationally significant because they can override intended governance constraints through untrusted inputs. When an agent retrieves and summarizes external content, malicious instructions embedded in those sources can induce the system to ignore policy, disclose restricted information, or generate

deceptive outputs. When the agent is connected to tools, the risk expands from wrongful text generation to wrongful action, including unauthorized dissemination, exfiltration of documents, or manipulation of outbound communications. These risks are foreseeable in any externally accessible system, and regulators should treat reasonable defenses as part of the issuer's governance obligations, not as optional quality enhancements. Where an issuer deploys an agent to interact with investors, it must be able to demonstrate that untrusted content cannot override system-level constraints, that tool permissions are tightly controlled, and that release mechanisms prevent the automated publication of market-sensitive claims without accountable review.

This framing matters because it preserves legal accountability while providing regulators with concrete supervisory levers. If a system can be induced to generate noncompliant statements through foreseeable attack paths, the deficiency is not merely technical; it is a failure of internal controls over external communications and sensitive information. In regulated markets, such controls are expected to function under realistic conditions, which include adversarial behavior. Accordingly, agentic deployments should be evaluated on whether they can be operated safely in the environments in which they are actually exposed, including public-facing channels, ambiguous user prompts, coordinated misinformation, and attempts to probe for confidential information.

## **Privacy, Personal Data, and the Architecture of Investor Interaction**

Investor-relations agents can process personal data even when the issuer's objective is merely to explain public information. Investor queries may include identifiers, account-related context, or behavioral patterns, and persistent conversation memory and personalization can amount to profiling or inference. Data protection regimes, particularly in the European context, therefore constrain not only how issuers store and reuse queries but whether certain personalization features can be justified at all without robust lawful bases and privacy-by-design measures. These constraints are not separate from market-integrity concerns. Systems that minimize identifiability by default reduce the risk of investor-specific informational advantages, narrow the attack surface for prompt-based leakage of conversation histories, and limit the incentive and capability to segment audiences in ways that can resemble selective access.

For regulatory oversight, the critical point is that system design determines both privacy compliance and the credibility of issuer assurances about equal access and controlled disclosure. Where personalization is offered, it should be structured to modify presentation, language, and explainability while preserving equality of underlying informational content. Where data is retained beyond short operational windows, the issuer should be able to demonstrate a documented lawful basis, clear limitations on purpose, restricted internal access to linkable records, and controls that prevent the system from regurgitating identifiers or prior conversations in response to manipulation attempts. A regulator assessing compliance should not be asked to accept generalized statements about anonymization or confidentiality; it should be able to examine architectural choices, access controls, and retention policies that render those statements verifiable.

## Regulatory Strategy and Supervisory Expectations

A risk-based supervisory approach should distinguish between systems that reorganize already-public information and systems that can create new market-moving statements or pathways to nonpublic information. This distinction should not be rhetorical; it should determine the issuer's required control environment and the regulator's supervisory posture. Where a system is designed to provide neutral explanations of public filings, the core expectation is that outputs remain within the public record, are accurately sourced, and are presented with appropriate limitations that prevent the agent from asserting as fact what it cannot verify. Where a system can generate interpretive claims about financial condition, performance, guidance, or market-sensitive events, or where it can access internal repositories or external sources of uncertain provenance, the system should be treated as a high-risk communications channel requiring disclosure-grade governance.

Regulators should state clearly that external-facing outputs generated by an issuer-deployed agent are corporate communications attributable to the issuer when delivered under the issuer's authority and on its channels. This does not require new substantive law; it operationalizes existing principles of issuer responsibility for statements disseminated to the market. The corollary expectation is that issuers must maintain internal controls capable of ensuring that agent outputs meet standards of accuracy, non-misleadingness, and market-integrity compliance under realistic operational conditions. Supervisors should evaluate not only what a system is intended to do but what it can do, because capability and exposure determine risk. An issuer that asserts its agent "only uses public information" should be expected to demonstrate enforceable boundaries in data ingestion, retrieval, and response generation, rather than relying on policy statements or user-facing disclaimers.

## Minimum Control Baselines for Investor-Facing Agentic AI

Regulators should require that issuers deploying investor-facing agents implement controls comparable in rigor to those used for other market communications, adapted to the distinctive risks of agentic behavior. The baseline should begin with governance ownership and documented accountability, including a defined control owner for the system, a pre-deployment risk classification tied to the system's capabilities and accessible data sources, and a change-management process that treats model updates, retrieval-source changes, and tool integrations as controlled modifications requiring review. These expectations are enforceable because they are auditable. They allow supervisors to examine whether an issuer has identified the system as a regulated communications process and assigned it to accountable persons with authority to approve, constrain, and suspend it.

Where the agent produces outputs that could be interpreted as factual assertions or interpretations about financial results, guidance, material events, or other market-sensitive matters, regulators should expect a human release gate that is meaningful rather than nominal. The operational requirement is that such outputs are not delivered to external users unless reviewed and approved by responsible personnel under documented procedures. In addition, issuers should retain a supervisory record sufficient to reconstruct what the system did at the time of communication, including the inputs received, the sources retrieved, the prompts and configuration parameters in effect, and the final output delivered. This is not recordkeeping for its

own sake; it is the mechanism by which regulators can assess whether failures were foreseeable, whether controls were bypassed, and whether the issuer maintained a defensible control environment comparable to other disclosure processes. Data segregation is central to making “public-only” claims credible. External-facing agents should not have technical reach into confidential or restricted repositories unless the issuer can demonstrate a narrowly scoped, controlled, and monitored retrieval design that prevents inadvertent inclusion of inside information in outputs. Supervisors should treat mere policy prohibitions as insufficient where the system has access paths that could be exploited through ambiguous prompts, retrieval errors, or adversarial manipulation. The enforceable expectation is demonstrable: approved and enumerated knowledge sources, robust tagging and access restrictions at ingestion, and retrieval-layer constraints that prevent mixing of public and nonpublic materials in response generation for external audiences.

Tool-enabled autonomy should be treated as an escalation trigger for heightened requirements because it expands the system’s capacity to cause harm. An agent that can browse external content, query internal systems, send messages, or publish content should be subject to least-privilege access controls comparable to privileged-access regimes in cybersecurity governance. Strong authentication for tool invocation, separation between external interaction layers and internal systems, and sandboxing of permissible actions are not optional enhancements in this setting; they are necessary conditions for controlling foreseeable misuse. Where publishing or outbound dissemination is possible, regulators should expect that publication is constrained to controlled workflows with human authorization and that the system cannot autonomously post, file, or distribute market-relevant statements.

Adversarial resilience should be explicitly incorporated into supervisory expectations because external-facing systems will be tested. Issuers should be expected to implement reasonable defenses against prompt injection and related manipulation, including restricting the domains and repositories from which content can be retrieved, validating and filtering retrieved content, and preventing untrusted content from overriding system instructions. Supervisors should also expect periodic testing against foreseeable attacks, documented remediation, and an incident response capability tailored to AI-driven misinformation, leakage events, and integrity failures. These expectations are verifiable through evidence of testing protocols, logs, response playbooks, and the technical configuration of retrieval and instruction hierarchies.

## **Equal Access, Personalization, and the Prevention of Selective Disclosure**

Regulators should clarify that personalization in investor-facing agents is permissible only to the extent it does not create content privilege or differential access to market-relevant facts. The compliant use case is differential presentation of the same public information, such as plain-language explanations, language translation, accessibility improvements, and consistent navigation of public filings. The noncompliant risk arises when personalization changes substantive informational content, emphasis, timing, or implied certainty in ways that could advantage some investors over others or enable inference of nonpublic information. Supervisory guidance should therefore discourage issuer deployments that provide materially differentiated knowledge bases to selected audiences unless equivalent access is provided to the broader

market or the interaction is demonstrably constrained to non-material public information under enforceable controls.

Supervisors should also scrutinize the use of analytics derived from investor interactions. Sentiment monitoring and query analysis can be legitimate tools for improving clarity and identifying areas of confusion, but they can also facilitate selective narrative management or targeted engagement strategies that raise market-integrity concerns. The supervisory question is not whether analytics exist but how they are used and governed. Issuers should be expected to document the purposes for which interaction data is processed, limit internal access to aggregated insights where possible, and ensure that analytics do not become a mechanism for tailoring market-moving communications to particular audiences in a way that undermines equal access and fair disclosure principles.

## Conclusion

Agentic AI in investor relations is best understood as an extension of regulated corporate communications into a high-velocity, interactive, and adversarial domain. Existing disclosure, market-abuse, and antifraud frameworks already establish the substantive obligations that must govern issuer statements, the handling of inside information, and the preservation of market integrity. The regulatory task is to operationalize those obligations for systems that generate and distribute text dynamically, retrieve content from mixed-trust environments, and can be manipulated by hostile inputs. A coherent supervisory approach will preserve accountability by treating issuer-deployed agent outputs as financial communications, while requiring governance ownership, enforceable data segregation, human release gates for market-sensitive content, auditable records, least-privilege tool control, adversarial resilience testing, and privacy-by-design architectures that minimize identifiability and constrain profiling. Properly regulated, agentic AI can expand access to public information and improve investor comprehension without eroding the core principles of accurate, timely, and equal disclosure on which market confidence depends.