Please quote as: Schmitt, K. & Grauer, T. (2026). GENERATIVE ARTIFICIAL INTELLIGENCE GOVERNANCE: A STRUCTURAL PRACTICE PERSPECTIVE. European Conference on Information Systems (ECIS), Milan, Italy.

GENERATIVE ARTIFICIAL INTELLIGENCE GOVERNANCE: A STRUCTURAL PRACTICE PERSPECTIVE

Completed Research Paper

Kevin Schmitt, University of St. Gallen, Switzerland, kevin.schmitt@unisg.ch

Teresa Grauer, University of Kassel, Germany, teresa.grauer@uni-kassel.de

Abstract

We analyzed how eight organizations seek to govern Generative Artificial Intelligence (GenAI). We identified five key governance roles (i.e., technology owner, business owner, knowledge owner, risk and compliance, and end user). Moreover, we showcase that each role controls different resources that become critical at different lifecycle stages. Specifically, we demonstrate that the focal power dependence relation moves away from "the business depends on engineers to build AI" toward "engineers depend on knowledge owners, risk and compliance, and end users to steer AI." We further showcase three organizational coordination mechanisms (i.e., AI Business Unit, AI Squads, and AI Spine) and their respective effectiveness in managing evolving power-dependence relationships in an AI governance context.

Keywords: Artificial Intelligence Governance, Generative Artificial Intelligence, Artificial Intelligence Frontier, Power Dependence Relationship, Structural Artificial Intelligence Governance Practices

1 Introduction

AI governance can be understood as aligning AI with business objectives (i.e., capturing the upside) and legal requirements (i.e., avoiding the downside). Nonetheless, striking that balance is more complex than anticipated.

First, GenAI's status as a novel design material makes value creation harder than expected (Yildirim et al., 2023). Even as new GenAI models achieve ever higher scores on evaluation benchmarks, many organizations that have deployed GenAI report little or no tangible business impact (Warren, 2025). For example, in 2025, Australia's Commonwealth Bank announced plans to cut 45 customer-service jobs and replace them with an AI "voice bot." When the bot failed to match human performance and call volumes increased, the bank made a U-turn: it apologised to staff and rehired the employees (Chalmers, 2025).

Second, because GenAI is probabilistic and an ongoing innovation, it is more complex to control than non-probabilistic and static technology (Berente et al., 2021; Ji et al., 2022). For example, in mid-2024, Anthropic's Claude Opus 3 shipped with roughly 350 words of prompt-based instructions (i.e., system prompt); by May 2025, Opus 4 required more than 1,700 words (Anthropic, 2025). Each additional word tightens control. Still, the ballooning system prompts are themselves a symptom of the core problem. Controlling a generic, probabilistic, and ongoing evolving technology forces organizations into a perpetual race to extend guardrails as people discover ever-new, and sometimes risky, uses.

Prior research has begun to map the governance landscape for AI. Researchers have defined AI governance as a system of rules, practices, and tools that aligns AI use with organizational strategy, legal requirements, and ethical principles (Mäntymäki et al., 2022), and have drawn extensively on adjacent disciplines such as IT governance and data governance to do so. A growing stream of work identifies who should be involved in AI governance and why, offering initial accounts of key stakeholders and decision rights (Janssen et al., 2020; Papagiannidis et al., 2025; Shneiderman, 2020).

However, far less is known about how organizations should design and orchestrate these responsibilities in practice, and how the shifting power dynamics among roles shape governance effectiveness (Papagiannidis et al., 2025). This gap is particularly acute for GenAI, whose open-endedness and inscrutability make governance a moving target rather than a one-time configuration (Berente et al., 2021; Feuerriegel et al., 2024).

Drawing on the insights of eight Swiss financial-industry players, we develop a systems-theoretic framework for structural AI governance. Structural AI governance is the formal allocation and organizational coordination of responsibilities across roles. Our systems-theoretic approach aims to help practitioners understand how AI governance needs to evolve as GenAI use cases move from implementation to ongoing evolution. Specifically, we show that the critical resources, and thus the locus of power, shift from engineering capacity toward domain knowledge, risk & compliance interpretation, and end-user feedback. Lastly, we provide concrete guidance on designing organizational coordination mechanisms, roles, and their accompanying responsibilities, enabling effective structural AI governance.

2 Conceptual Background

2.1 Generative Artificial Intelligence as One of the Current AI Frontiers

Generative AI (GenAI) is a current frontier of AI, encompassing technologies that can autonomously generate (e.g., via prompt instructions) text, images, video, and audio outputs based on the data they were trained on (Feuerriegel et al., 2024). In that, Generative AI is characterized by two unique characteristics:

Open-endedness. GenAI uses flexible interfaces, most often natural language, and accepts and produces multiple modalities without fixed syntax. Probabilistic generation allows the same prompt to yield multiple plausible responses, expanding utility but reducing predictability (Feuerriegel et al., 2024). . GenAI can hallucinate, confidently producing inaccurate or nonsensical content (Ji et al., 2022). Also, open-ended interfaces invite prompt injection and other forms of misuse. In one case, a GenAI-powered Chevrolet dealership chatbot was manipulated into offering a car for $1 (Landymore, 2023).

Inscrutability. Training on vast corpora, often billions of data points, creates behaviors and edge cases that organizations struggle to anticipate (OpenAI, 2024; Sparks, 2024). Probabilistic outputs emerge from high-dimensional correlations rather than explicit, human-readable rules, which obscures how systems reach conclusions. Moreover, even with the source code and model weights, internal representations remain difficult to interpret. (Berente et al., 2021; Feuerriegel et al., 2024). This inscrutability carries operational consequences. For example, clinicians hesitate to rely on GenAI for diagnoses and treatment plans without transparent reasoning. (Chen et al., 2024).

2.2 Artificial Intelligence Governance

AI governance is a collective term. Also, the AI governance literature is primarily built on research in adjacent technology governance disciplines (e.g., IT governance and data governance) (Mäntymäki et al., 2022). Accordingly, the literature tends to focus on defining AI governance, comparing it with other governance disciplines, and exploring how it differs from them.

AI governance is commonly defined as *"a system of rules, practices, processes, and technological tools that are employed to ensure an organization's use of AI technologies aligns with the organization's strategies, objectives, and values; fulfils legal requirements; and meets principles of ethical AI followed by the organization." (Mäntymäki et al., 2022, p. 604)*.

Overall, the literature concludes that governing AI is more complex than governing other information technologies. This is because AI's unique characteristics (Berente et al., 2021) make it difficult to anticipate how AI systems will behave. However, little is known about the mechanisms (e.g., structural governance mechanisms) that organizations use to govern AI and keep it aligned with their governance goals.

To illustrate the challenges of AI governance, it is essential to recognize that AI governance depends on two conditions: (1) the ability of the controller to direct the controlee in achieving a desired objective (i.e., the Directivity Condition) (2) the existence of a clearly defined objective (i.e., the End-State Condition) that can be communicated by the controller to the controlee.

First, as AI becomes more inscrutable, adaptive, and autonomous, achieving even a well-defined end state by following specific guidelines (i.e., Directivity Condition) is far from guaranteed. Unlike traditional software, where behavior can be explicitly hard-coded, AI systems are statistical models that may produce varying outputs even when the inputs remain identical (Berente et al., 2021; Feuerriegel et al., 2024; Ji et al., 2022). The years-long, repeatedly delayed move toward fully autonomous driving, thwarted by innumerable real-world edge cases and underscored by the 2023 Cruise robotaxi pedestrian-dragging incident, illustrates how brittle statistical systems can be in open environments (Roy, 2024).

Second, it is difficult for the controller to articulate a desired end state (i.e., End-state Condition) when they are permanently confronted with AI capabilities that expand both the scale and scope of what is possible (i.e., AI as an ongoing innovation). As a result, there is rarely consensus within or outside the organization on whether, or in what form, these new capabilities ought to be pursued. For example, Meta's AI chatbot, which at times allows "sensual" exchanges with teenagers, shows how rapidly evolving AI capabilities can outpace a shared sense of what is acceptable or lawful (Horwitz, 2025).

2.2.1 A Structural Practice Lens on AI Governance

Organizations use structural AI governance practices to specify who makes decisions (Janssen et al., 2020; Papagiannidis et al., 2025; Shneiderman, 2020) These practices also clarify the rights and responsibilities of individuals and user groups within the organization (Papagiannidis et al., 2025; Salunke et al., 2011). Moreover, organizations formally document these rights and responsibilities, drafting, reviewing, approving, and updating them in line with organizational values, regulatory requirements, and ethical guidelines (Papagiannidis et al., 2025; Too & Weaver, 2014).

Despite a growing body of research on structural AI governance practices, much of the work still centers on who should be involved and why. These insights provide researchers and practitioners with an initial understanding of key stakeholders in the context of AI governance. However, far less is known about how organizations should design and orchestrate vertical and horizontal decision rights, and how power dynamics shape the effectiveness of these rights. Yet such insights are critical for understanding how organizations must be designed to support the effective exercise of rights and responsibilities, rather than merely defining them.

3 Research Method

3.1 Case Study

Given the limited existing theory and evidence related to our research question, we adopted an inductive theory-building approach using multiple cases (Eisenhardt, 1989). This approach is particularly suitable when data exhibit variance, as is the case in our study. Furthermore, multiple cases are often more effective than single cases for theory development, as replication across cases enhances the accuracy, parsimony, and possible future generalizability of the resulting theory (Eisenhardt, 1989).

Our embedded design focuses on individuals involved in AI governance as the unit of analysis. By maintaining this focus, we aim to enhance the richness and precision of the theory development process (Eisenhardt, 1989; Siggelkow, 2007). Additionally, this focused approach increases the likelihood that the developed theory will be tested in future research.

Our sample consists of eight organizations: three of Switzerland's largest banks, three prominent health insurers, one multinational financial services group, and a medical coding company. Three common patterns emerged for structural AI governance practices, suggesting that this sample size approaches theoretical saturation (Eisenhardt, 1989; Yin, 2017). All organizations had prior experience with AI.

However, their previous AI solutions had lower levels of autonomy, inscrutability, and learning ability. Also, these AI solutions were mainly isolated and often served a very narrow use case.

3.2 Data Collection Process

We conducted 61 semi-structured interviews with 45 representatives from my eight case companies. We primarily relied on interviews as a data source because they allowed us to capture employees' lived experiences at our eight case companies. Our choice of interviews as the primary method aligns with Walsham's (1995) argument that interviews provide the best access to understanding participants' experiences of lived events. The interviews lasted approximately 60 minutes each. Following established qualitative research standards, we conducted interviews until saturation, at which point themes began to recur (Myers & Newman, 2007). Thus, we carried out 61 interviews between March 2023 and March 2025 (Appendix A).

3.3 Data Analysis Process

We began by analyzing real-world data to identify a phenomenon relevant to both practitioners and academics. Thus, we first focused on identifying practical problems and then used them to challenge existing ways of thinking in the academic literature. After identifying a topic of interest (i.e., structural AI governance practices), we sought a sound theory to guide our analysis (i.e., power-dependence theory). After that, we began the "within-case analysis" and sought to understand the power and dependence composition of our case organizations, respectively. This was done inductively.

Next, we attempted to understand how the identified power and dependence mechanisms influence the structural governance of AI along the use case lifecycle (i.e., implementation, evolution, and abandonment). Again, we utilized an inductive approach for this analysis step.

In the subsequent phase, termed "cross-case analysis," we outlined causal templates across the case companies, illustrating the outcomes of different structural GenAI governance practices. Our approach is built on in-depth knowledge of the individual case companies.

3.3.1 Within-case-analysis

Codes serve as symbolic representations of descriptive or inferential data collected during a study (Myers & Newman, 2007). These codes facilitate the retrieval, categorization, and triangulation of data from interview transcripts by grouping elements associated with specific constructs or themes.

Starting with the within-case analysis, we employed a multi-stage coding scheme (Gioia et al., 2013). This approach enabled us to discern a data structure for all eight case companies, illustrating the relationships between first-order concepts, second-order themes, and aggregate dimensions. We began coding the data collected for the eight cases, focusing on first-order concepts, specifically power and dependence indicators along the GenAI use case lifecycle. We prioritized maintaining the participants' language and expressions to capture their true intent. If there were any ambiguities in the data, we contacted the interviewees by email or phone to seek clarification.

This approach enabled us to draw connections between 1st order concepts, elevating them to the status of second-order themes. Ultimately, we merged the second-order themes into three aggregate dimensions. Thus, we created our final data structure for our within-case analysis.

3.3.2 Cross-case-analysis

We started with the cross-case analysis to continue our analysis process and identify recurring patterns across the case companies. Accordingly, we systematically compared, contrasted, and triangulated the data across the case companies. The first-order concepts encompass power and dependency mechanisms across our case companies. The second-order themes represent organizational coordination mechanisms within these organizations. Furthermore, the third-order aggregate dimensions represent the various

structural AI governance practices envisioned by our case organizations. Moreover, our data structure comprises first-order concepts, three second-order themes, and three aggregate dimensions.

4 Results

We identified three structural AI governance coordination mechanisms (i.e., AI Business Unit, AI Squads, and AI Spine). Moreover, we identified five key roles involved in AI governance (i.e., tech owner, business owner, knowledge owner, risk and compliance, and end-user) and their positioning within the organization. Second, we outline the circumstances under which the three described structural AI governance coordination mechanisms are applicable and when they reach their limits, necessitating a change (see Figure 1).

4.1 The AI Spine

To govern frontier-level AI, such as GenAI, at scale, two of our case organizations adopted an AI Spine structural governance practice. Moreover, we use the AI Spine as an umbrella term to capture related designs, such as a Value Stream or Flat Organization, that may differ in labels but share a common logic. They centralize technology-related resources while partially centralizing business-related resources to ensure sustained, cross-functional governance. Moreover, the AI Spine runs above the business units, orchestrating five core roles embedded in each AI use case and enabling vertical and horizontal coordination at scale (see Figure 3).

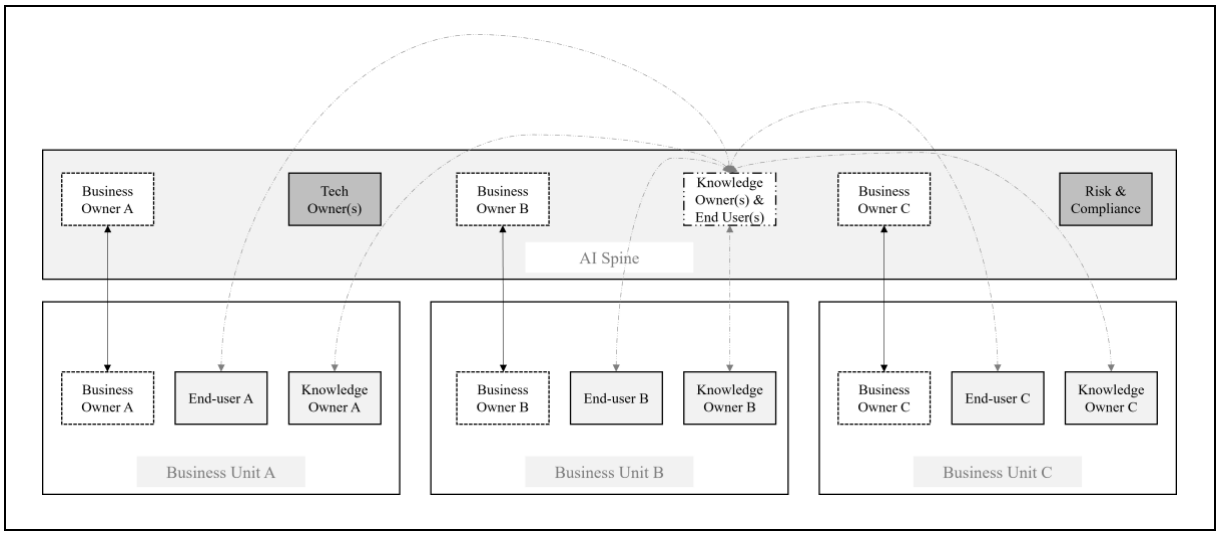


Figure 1. The AI Spine

4.1.1 Roles that Make it Work

The AI Spine centralizes the business knowledge required to rewire and continuously improve a firm's operations end-to-end, rather than serving as a distributor of technological knowledge to business units.

The Technology Owner is responsible for preventing the fragmentation of data flows and tools, which create inefficiencies and yield suboptimal performance across use cases. They typically oversee the creation of centralized data platforms, prompt libraries, models, evaluation, and technical performance metrics. This reduces rework and lowers marginal costs, helping to diffuse technical breakthroughs across the organization.

The Risk & Compliance role is a permanent fixture of the AI Spine to ensure compliance issues are managed throughout each iteration of a use case. The permanence of this role also supports learning and organizational memory, meaning that new compliance issues, once addressed, can be addressed more quickly in other use cases.

Acting as a bridge between the AI Spine and business units, the Business Owner enlists representatives from each unit to ensure solutions deliver real value to the business. They identify end-to-end processes that are ideal candidates for AI, helping to determine nontechnical baselines and targets. They are also responsible for ensuring that proven use cases are implemented across units for the identified processes, when applicable, while under-delivering use cases are dropped. Without this critical role, proofs of concept fail to gain momentum beyond the minimal viable product stage or remain confined to a specific business unit.

Working alongside the Technological Owner, the Knowledge Owner ensures that both explicit and tacit knowledge are captured in the technology. They curate ground truth and provide important feedback on AI components, thereby becoming de facto repositories of organizational knowledge. For example, prompt libraries need to be up to date with company policies and safety guardrails to generate appropriate responses. This role quickly identifies issues, reducing rework and costly escalations as GenAI use cases are implemented and evolve. They also help to rein in AI work when results are strong enough from a business perspective, regardless of technical performance.

Finally, the End User represents the primary users of AI solutions. Underperforming use cases lead to low adoption rates, workarounds, and, in some cases, Shadow AI. This role exists to validate that a use case will effectively help those who will be using the technology. Beyond pre-implementation work, this role also provides critical feedback on post-implementation performance from a user standpoint (e.g., to identify edge cases), helping drive future improvements and, in some cases, providing data (e.g., via feedback, edits, and ratings) to improve the solution in the next iteration.

4.1.2 The AI Spine as a Structural Coordination Mechanism

Among our case companies, two organizations created an AI Spine. Moreover, building an AI Spine requires substantial upfront investments in talent and organizational design, as well as a commitment to playing the long game.

4.1.3 Building an AI Spine

Our case organizations use three rules to bring the AI Spine to life. Specifically, they follow the mantra protect, connect, and improve. Two running examples illustrate the benefits of this approach. A retail bank's AI Spine spearheaded the implementation of a GenAI email assistant. At a medical coding company, GenAI is used to automate medical billing coding (see Table 1).

GenAI Email Assistant	GenAI Medical Coding
• Retrieves relevant information • Automatically drafts customer reply emails • Improves response speed and consistency • Allows human review and editing before sending	• Reads medical documents • Identifies diagnoses and medical procedures • Converts clinical information into standardized medical billing codes from the International Classification of Diseases • Reduces manual coding effort and errors

Table 1. Explanation of the Running Examples

#1 Rule: Protect. When the bank initially built the email assistant, the efforts were primarily technological, focusing on integrating fragmented data sources (e.g., the company website and internal data). While technically impressive, feedback from customer service representatives was underwhelming. Generated responses were plausible but often incomplete or subtly misaligned with the bank's standards, and the messages' tone was inconsistent, requiring multiple modifications before they were deemed fit to be sent back to customers without incurring potential liability. These results underscored the importance of tacit knowledge (i.e., knowledge not encoded) embedded in the agents' work, which had not been captured by scraping available data. The AI Spine built on this feedback to revise prompts, changing how company data was used to generate responses and steer GenAI model's behavior. Moreover, the realization that especially not encoded knowledge was necessary to protect the organization, the AI Spine to veto underperforming releases while advancing when all criteria were met.

> *"And it is being listened to [Our opinion] very well. This is the story about the veto […]. We took it very seriously. And overall, front-line employees (i.e., those in the customer service center) are represented there [AI Spine Meetings]." [3.5.2_RBC. Department Head Customer Service Centre]*

The AI Spine also sought to make tacit knowledge more explicit. Customer service employees were compensated for providing feedback on the assistant, including ratings of 1-5 for generated responses, while manual edits to those responses were logged for review. This provided critical data to evaluate the email assistant's performance, creating a feedback loop that surfaced the organization's unwritten norms on tone, intent, escalation thresholds, and edge cases that can take a significant amount of time to resolve. Since then, the AI Spine has routinized the use of ratings and the logging of edits to AI-generated content as key metrics for other use cases, complementing technical metrics. This allows them to identify cases where efforts to address model, prompt, and data shortcomings are likely to generate value, focusing human effort where it will be most valuable, reducing employee turnover, and alleviating initial concerns about job replacement.

> *"We saw that an uncurated knowledge base does not have the quality you need to achieve good results. […] Because we cannot do it ourselves, we do not have the resources, meaning not enough people to build this knowledge base at all. And we also do not have the domain knowledge. […] And then the next step was to motivate the people who would benefit most from it, in this case, the customer service center team. […] Within a relatively short time, we were able to create more than 800 entries with the help of these people […] and we benefited from employees' experience as well. They said: this is a question that comes up often, and this is how I answer it. […] If, during development, I change one or two words in a prompt, the output is sometimes completely different. And that is why we also developed testing pipelines where we have around 50 ideal questions and ideal answers. […] I have my biased view, I am sitting there coding and I have my 50 static test cases, but what it looks like in production and whether this change is actually usable still has to be judged by the end users. And only once I have that "OK" do I know: okay, this works, we can put it into production. […] Because otherwise it would be too complex. We would have too many test cases and it simply would not be manageable. […] It is important to mention that we log all requests. […] And each answer must, or ideally will, be rated by a human, i.e., by the user of the tool, on a scale from 1 to 5, and additionally with a feedback text. That helps us in ongoing evaluation […]." [3.13_RBC. AI Engineers]*

Efforts at the medical coding company were also initially met with underwhelming technical performance. To address this issue, AI engineers argued for additional data collection, aligning with the standard view that more data leads to better performance. However, nontechnical roles in the AI Spine reviewed the results and challenged this argument. On closer inspection, it was found that "ground-truth" labels used to train the model were inconsistent, and that the system was, in fact, outperforming human coders. The issue was that human coders had mislabelled some of the data.

> *At the beginning […] we made some major mistakes. We believed the medical coders. They said, "[Name of the Person], do not worry, it is really easy. Everything is black and white. There is this thick handbook that explains how it is. And the catalogs specify how the codes must be used. It is all very black and white, everyone codes the same way." It turns out that is not true at all. Everyone interprets it and sees it differently. The differences are very large. And there are very few quality controls. [8.1_ID. Chief Executive Officer]*

The ability to discuss both technical and business performance evaluation elements, including accuracy, error types, and ground truth, using a common language reflecting the multifaceted nature of GenAI performance, helped the organization implement the proper fix (i.e., curating better ground truth data) for the next iteration and avoid wasting time and resources training a model on incorrect data.

#2 Rule: Connect. Both the bank and the medical coding company realized that business units were reluctant to fund GenAI use cases due to the high degree of uncertainty surrounding these initiatives, stemming from the open-ended nature and the technology's inscrutability. Ideas can appear promising

and sometimes work well on a small scale or as proofs of concept, but industrializing them and extending their reach across processes would prove costly and challenging. Both firms set out to define explicit GenAI goals in their corporate strategies, sending a strong signal across units that GenAI was now a business priority. Establishing the AI Spine as a separate entity with a clear connection to executives helped to maintain strategic alignment. However, having its own funding seeded by executives and financed by business units allowed the Spine in both firms to maintain some degree of independence. At the bank, the AI Spine funds use cases across multiple phases through micro-grants tied to the achievement of objectives and key results, covering both technical and nontechnical criteria, and a similar pattern was observed at the medical coding company. This approach helps build solutions and manage risk incrementally for long-term initiatives. In both firms, the Spine acts as a catalyst for achievements and learnings that can benefit all business units. While removing funding for GenAI initiatives from business units was initially perceived as problematic, over time, it encouraged bolder innovation and a healthy degree of competition between units. Based on these insights, it became clear that the technology itself is not decisive, but the organizational design is.

> *"I think the algorithm part is not defensible. […] But I think the learning process and clarifying the roles and how they interact is a competitive advantage. […] It is really anchored in the strategy […] so that this becomes a scalable model for us. […] You start somewhere, right, […] and then you realize the setup does not work […] and you say: no, we cannot keep solving this temporarily with just a few motivated people, otherwise it will not scale, […] so what does this mean for our setup and our governance? […] That is one of the reasons why we tried to focus […] Otherwise, we simply do not have the capacity to think through each time where exactly the boundaries are, right? […] In the end, it is not predictable how quickly things will develop. We have to set ourselves up so that we can cope with it." [3.11_RBC. Deputy Chief Operating Officer]*

Members of the AI Spine in each firm also realized that process knowledge was highly fragmented. Each unit had a relatively clear idea of how its part of a given process worked, but a clear vision of the entire process end-to-end across units was missing. The AI Spine championed the mapping of key processes where GenAI use cases had been proposed, involving stakeholders from each business unit, until these processes could be drawn accurately. With these representations, they could clearly communicate the scope of GenAI use cases, their potential for expansion, and their applicability across multiple use cases, as illustrated by the creation of a new revenue stream for the medical coding company. Unlike AI centers of excellence that are primarily geared toward centralizing technological know-how, the AI Spine also centralizes knowledge of business processes to fuel innovation.

Connecting with strategy and adopting an end-to-end process view only yields results if there are forums dedicated to disseminating information from the units to the Spine and back. Adopting proven practices from scaled agile software development, GenAI leaders leverage the Spine as an authoritative platform for regular, short stand-up meetings. At the bank, for instance, AI spine members share lessons learned, discuss ongoing and future priorities, and communicate changes to knowledge repositories, such as prompt libraries, during these meetings. A "GenAI Summit" event is also held every month and includes end users to keep them updated with progress, show them demos, and gather their feedback. This lowers the risk of failure by making necessary adjustments during the development process rather than through post-go-live rework, and it also helps to manage the expectations and the apprehensions of future users.

> *"Since we have been working in value streams [i.e., this paper calls it AI Spine], which focus on topic blocks and include everyone, such as product owners, business engineers, people from operations, and we as primary stakeholders, the value streams as a whole remain responsible for operations. […] Every two weeks we have these reviews where we can dial in, and the streams present their results and at the same time show the plan for the next two weeks. […] There are sometimes 30 to 40 people in these reviews. […] That means we now know much more precisely what is going on in the bank overall. We can contribute in these calls and share our opinions, and they are very well heard. […] Without this, it would be much harder to understand where the development stands. Also, I think, the organization has realized that with GenAI there is much more input needed from business. It is more like a work and progress, and*

you need to find a way to bring everyone together. This was just not possible with our traditional set-up" [3.5.2_RBC. Department Head Customer Service Centre]

#3 Rule: Improve. At the medical coding company, early GenAI trials appeared promising from a technical standpoint but still did not deliver sufficient performance to be viable in production. As the AI Spine pushed to analyze the root cause of this underwhelming performance, teams found that medical coding knowledge was so specialized that it was underrepresented in the training data of the GenAI. Yet the GenAI's understanding of the natural-language portion of the use case was strong, so the organization opted to fine-tune it with medical coding data. Although this helped improve performance, the results still fell short for production use. It took further iterative experimentation to get to a viable solution: using four GenAI models to code medical reports. Two models were tasked with coding the primary diagnoses, while another two coded the secondary diagnoses. This distributed approach improved performance for the original medical coding use case. Moreover, it led to the development of an ecosystem of GenAI models, each with its own specialty (e.g., clinic- or code-specific), expanding coverage to handle more code types and multiple languages. The company's AI Spine's iterative approach to experimentation helped improve the organization's GenAI-driven value by taking calculated risks and incorporating learnings from previous iterations.

> *"We are now testing the GenAI and looking at how high the performance percentages are. [...] We have [...] chained several models one after the other. That means we learned to not give everything to one model [...] and in this way the overall result became the most accurate. [...] This now raises a lot of very interesting questions for us for the next phase [...] Do we keep these four GenAI's as they are, [...] or does it make sense to say, okay, we build one model or a combination of models that only do orthopedics and others that only do births? And that is where it will be interesting, I think, [...] how broad or how specific these GenAI's need to be. [...] I think, to measure which combination of models is necessary. [...] However, this also means that we continuously have to pay for training GenAI's in order to keep testing new hypotheses." [8.1_ID. Chief Executive Officer]*

The bank also capitalized on its AI Spine's approach to iterative experimentation, treating GenAI as a constant work in progress. While the GenAI email assistant eventually led to significant time savings for customer service agents (i.e., 15 percent time reduction), it took multiple small-increment rollouts that continually raised the bar to achieve them. Having clear technical and business performance targets at each iteration defined the goal of whether the solution would be rolled out to end users, and the AI Spine oversaw ensuring that the technological infrastructure they were building would support this iterative, incremental approach (e.g., Provisioned Throughput Unit that you allocate to a provisioned deployment so you can achieve a target, predictable throughput for prompts and completions). It also gave teams time to learn progressively and to assimilate new technological developments in a field that evolves at an unrelenting pace. For example, embedding varying degrees of work automation led to a more granular view of performance: one version of the assistant would work well for verifying messages written by humans but would perform poorly at writing messages, driving work for the next iteration. Thus, via the AI Spine, the organization enabled itself to treat AI-related work as an ongoing and iterative process.

> *"However, with AI, improvements can still be made to the same functionality even a year later […]. The AI model must continue to be improved […] otherwise, one might think that it can simply be left as is. […] one must be aware that it is an iterative, continuous process." [3.3.1_RBC. AI Engineer]*

Keeping the momentum to continue innovating with GenAI is a challenge that requires a delicate balance between exploiting existing use cases and exploring new ones. At the bank, identifying new GenAI opportunities proved more challenging than expected once initial use cases started to diffuse across processes. Previous applications of AI (e.g., machine learning for fraud detection) had a narrow focus. They could only be transposed across contexts with very similar conditions. The open-ended nature and the rapid evolution of GenAI constantly reconfigure the landscape of potential applications. The AI Spine identified a critical need to understand better how employees interact with GenAI across

processes. By studying these interactions, running employee surveys and shadowing sessions, they were able to identify eight new potential use cases, two of which have since been deployed alongside of the email assistant: an AskHR chatbot (i.e., answering HR-related questions such as employee benefits), and an Employee Handbook chatbot (i.e., answering employee questions as to what is permitted and forbidden in the execution of their job).

4.1.4 When to Apply and When to Switch

The AI Spine structural governance practice is most effective when an organization must govern a broad portfolio of GenAI use cases spanning multiple business units. However, the added coordination and overhead rarely pay off when (1) the organization relies mainly on well-established AI solutions; (2) frontier AI applications are limited to a few narrow pilots; or (3) most AI use cases are procured and governed by external providers. In these circumstances, an AI Business Unit or AI Squads (see below for more information) structural AI governance practice is typically more sustainable.

4.2 The Journey to Building an AI Spine

Among our case organizations, only two build an AI Spine. This is not surprising, as the AI Spine requires significant time, effort, and financial resources. Without a clear strategic direction and commitment from executives, this structure can quickly become a burden. The other organizations that were part of our research adopted two alternative structures that we see as gradual steps in their AI journey.

4.2.1 The AI Business Unit

Four of our case organizations began their structural AI governance journey by following an AI Business Unit approach. This umbrella term encompasses different centralized setups, such as a Digital Transformation Unit or an AI Center of Excellence, that vary in name but share a common logic. They primarily centralize technology-related resources.

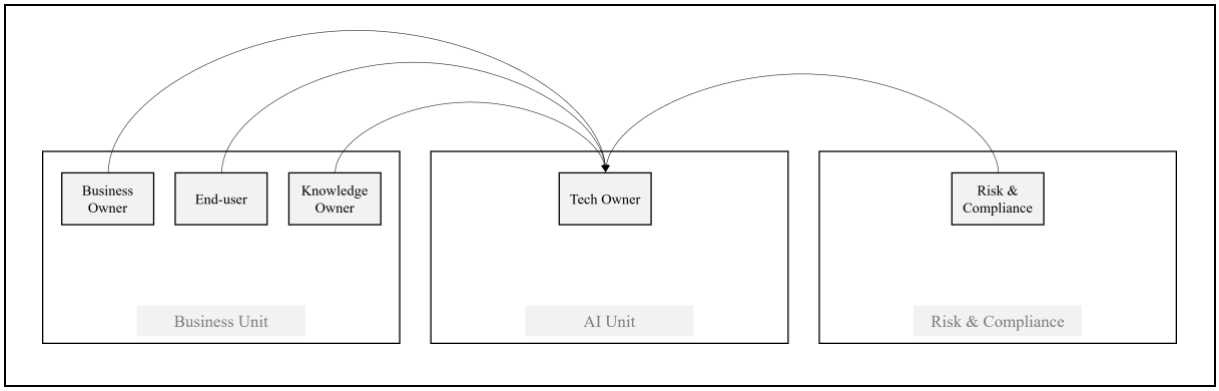


Figure 2. The AI Business Unit

However, the very structure that speeds the first wins can soon become a governance bottleneck. Data, domain expertise, and compliance reviews often remain in other business units, each with its own priorities, so the AI Unit can end up spending more time negotiating access to these resources than building GenAI use cases.

> *"With a simple chatbot, you can achieve results relatively quickly, but when introducing Conversational AI [he is talking about Large Language Models], there are many dependencies, at least for us at [Company RBB]. One should be aware of these dependencies to avoid showstoppers." [2.1_RBB. Lead Change Management]*

Concentrating technical talent in an AI Business Unit jump-starts AI initiatives, but as the initiatives mature, the locus of value generation shifts to knowledge owners, business leaders, and end users. Because the technology is open-ended in scope and opaque in operation, their iterative guidance is

essential to align outputs with business objectives. However, in the absence of effective horizontal and vertical orchestration mechanisms, employees do not feel obligated to contribute to governance efforts. Consequently, pilots proliferate while few scale, and underperformers persist.

At a Swiss health insurer, a GenAI project faltered because leaders ran it like a conventional IT project. The technology team, seasoned in traditional IT systems, underestimated the sustained, cross-functional effort required to govern GenAI across its lifecycle. Governance was treated as an IT responsibility and was concentrated in the technology owner after deployment; data and business owners, and end-users lacked decision rights and the capacity to contribute meaningfully thereafter. A set-up that worked well in the past but proved fatal in the context of GenAI. Instead of downgrading the scope and scale of the GenAI use case, the tech owner attempted to compensate by absorbing the additional responsibility, but without a cross-functional coordination mechanism to access the necessary resources, governance became ineffective. Technical expertise was abundant; integrative governance was not.

> *"That is why it is funny that this is coming from me [Head of Digital Transformation], but that is actually the crucial point with all of these matters at the moment. We do not have a technological problem, interestingly enough. […] what we have are organizational problems. The question is how we can coordinate people to work on these artificial intelligence projects. These are not information technology projects. We need much more input from the subject-matter experts, because we do not understand the context of the data. The exchange of knowledge and perspectives would need to be much stronger, and the responsibilities would need to be distributed […]." [5.2_IB. Head of Digital Transformation]*

The result was predictable and public. On the insurer's website, the chatbot, among other things, replied "Congratulations" when a user typed "I have cancer." As a result, the insurer was left with a "zombie" GenAI, deployed but effectively ungoverned, with no plan for improvement.

4.2.2 When to Apply and When to Switch

Across our case organizations, the AI Business Unit approach proved effective when the AI solutions in question were not at the frontier, and various good practices were readily accessible to the organization. Typical examples included optical character recognition, simple AI models for email routing, and decision-tree chatbots. Because these systems exhibit limited autonomy, learning, and inscrutability, existing governance structures were sufficient. By contrast, once organizations adopted more sophisticated, frontier AI solutions, such as GenAI, they had to adapt their governance practices.

4.2.3 The AI Squads

Two organizations embedded small, cross-functional AI Squads to bring the tech owner closer to the business owner, the knowledge owner, and the end user, enabling these roles to collaborate more seamlessly. Moreover, the term AI Squads serves as an umbrella for related organizational designs, such as Cross-functional Teams or Agile Teams, that may differ in label but share a common logic: they concentrate technology-related resources and position them as close as possible to a specific business unit, without those resources fully becoming part of the business unit itself.

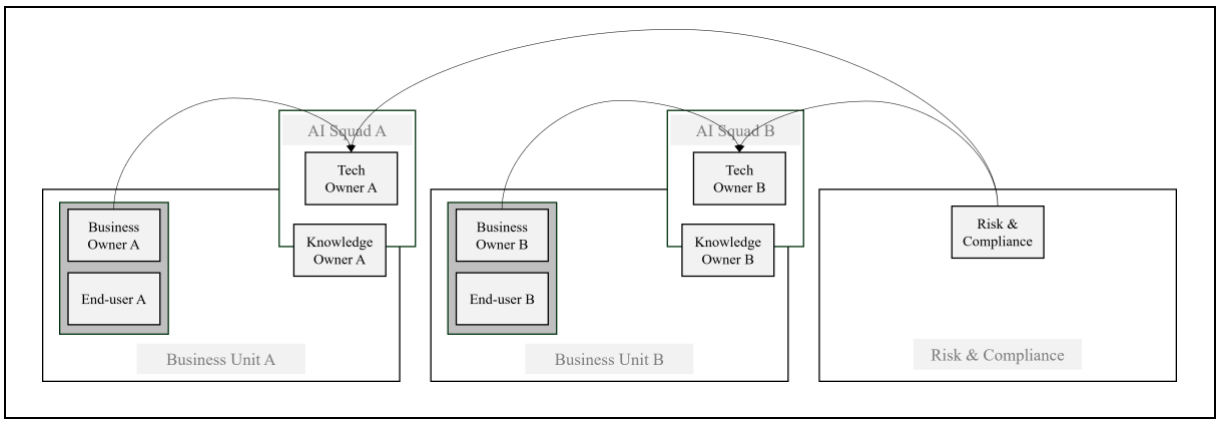


Figure 3. The AI Squads

Closeness does not solve the problem when incentives remain misaligned. However, this approach makes governance in the beginning phases of an AI use case lifecycle more effective. Nonetheless, maintaining high governance intensity after implementing the AI use case remains challenging. Once an AI use case goes live, the Tech Owner needs fresh data, may need to conduct periodic prompt updates, policy updates, and needs compliance sign-offs. AI Squads are expected to coordinate patching, data refreshes, prompt and policy revisions, and continuous monitoring. Yet outside the tech and knowledge owner, few roles have dedicated capacity or incentives, and the organizational orchestration mechanism (i.e., AI Squads) does not fully support the inherently iterative, ongoing work that frontier-level AI requires. For example, employees in business units are still evaluated on the business unit's yearly performance, rather than on how meaningful their input to the AI project was. As a consequence, accountability diffuses, while tech and data owners are pulled into a perpetual cycle of maintenance. The result: stalled innovation and a pipeline of GenAI use cases that fail to meet strategic objectives.

A Swiss insurer's chatbot shows how quickly things unravel when organizations only half-adapt to GenAI's quirks. The organization launched a customer-facing chatbot for insurance product information in record time. However, only a couple of weeks later, curious customers prompted it to recommend pizza recipes.

> *"We worked very closely with the business unit […] to capture all the points you need, all the things that have to be asked. […] As for maintenance, we are the ones who handle that. We just keep running the GenAI service, and the business units come to us when, for example, they have regulatory changes. […] The main difficulties, I would say, are implementing things with the existing IT resources and environments and coordinating with the business units. […] It is not that they actively work on it, they give input and we provide proposals. […] To do more, we would need more resources and people would take an active part in the development and governance of GenAI solutions. Otherwise, we have another failure like a chatbot […] At the moment, we reject most use cases because we do not have the resources to manage them." [6.3_IC. Conversational Marketing Manager]*

The tech owner quickly deployed a patch to update the guardrails, but the incident revealed just how much maintenance work is required to capture meaningful value from GenAI. Implementation had been straightforward; governance, however, proved unexpectedly demanding. As the GenAI portfolio expanded, the mounting maintenance burden (e.g., data source and prompt updates to address edge cases) stretched the Tech Owner to its limits, leaving little capacity to pursue new GenAI use cases. The "you build it, you run it" approach failed the organization in integrating a larger number of GenAI use cases.

Typically, the knowledge owner is a dedicated role. End users stay engaged when GenAI helps employees meet their key performance indicators. Hence, this setup enables a more iterative approach to AI governance, giving people with specific domain knowledge the opportunity to provide feedback and help the tech owner understand what GenAI outputs ought to look like (i.e., an orchestration-focused AI governance approach). Still, business owners and risk and compliance teams often disengage, treating governance as a one-off rather than an iterative, ongoing process. Because more stakeholders remain involved and most post-launch work is data-focused, risk and compliance can be managed periodically, allowing the organization to govern a small set of narrowly defined GenAI use cases effectively.

4.2.4 When to Apply and When to Switch

Across our case organizations, the structural AI Squads governance practice worked well when the GenAI portfolio was small and confined to a single business unit. However, as portfolios expanded, and especially when use cases spanned multiple business units, the limits of this structural AI governance practice became apparent, triggering the need for a new structural AI governance practice.

5 Discussion

Our analysis of eight organizations governing GenAI reveals that structural AI governance is not a one-time configuration but a dynamic organizational capability that must evolve as GenAI use cases move from implementation to ongoing operation.

First, governance consistently broke down not because organizations lacked technical capacity, but because the roles controlling the resources most critical to steering GenAI (e.g., domain knowledge, compliance interpretation, and situated user feedback) lacked the decision rights, dedicated capacity, and institutionalized routines to exercise meaningful influence after deployment. Second, the five roles we identified are not equally important at all stages; their relative criticality shifts across the lifecycle. Third, the three coordination mechanisms we observed (i.e., the AI Business Unit, AI Squads, and the AI Spine) are not inherently superior or inferior to one another. Their effectiveness depends on the scale and cross-functional complexity of the GenAI portfolio they are designed to govern.

5.1 Theoretical Implications

This paper makes two theoretical contributions to the structural AI governance literature. First, we extend existing structural AI governance research by moving beyond the question of who should be involved in governing AI toward the question of how governance responsibilities must be designed and orchestrated as GenAI use cases evolve. Prior research has largely treated governance roles and decision rights as static configurations (Papagiannidis et al., 2025; Janssen et al., 2020). We show that the critical resources, and therefore the locus of governance authority, shift across the lifecycle, moving away from engineering capacity and toward domain knowledge, risk and compliance interpretation, and end-user feedback as use cases move from implementation to ongoing evolution.

Second, we contribute a systems-theoretic framework that links five governance roles to three structural coordination mechanisms (i.e., the AI Business Unit, AI Squads, and the AI Spine) and specifies the conditions under which each mechanism is effective and when it reaches its limits. This provides structural AI governance research with a dynamic, lifecycle-sensitive account of how organizational design shapes governance effectiveness.

5.2 Practical Implications

At the role level, we give practitioners a concrete map of five governance roles, the resources each controls, and when those resources become the binding constraint on whether a GenAI use case delivers value without producing unacceptable outcomes. This helps organizations recognize when governance is silently failing, not because the technology underperforms, but because the wrong roles hold decision authority at the wrong stage.

At the structural level, we provide actionable guidance on when to adopt an AI Business Unit, AI Squads, or AI Spine, and critically, what signals should trigger a transition between them. Rather than prescribing a single governance blueprint, we equip organizations to match their governance structure to the cross-functional complexity of their GenAI portfolio and to adapt that structure as both the portfolio and the technology evolve.

5.3 Limitations and Future Research Directions

Our study has certain limitations that also offer opportunities for future research. First, we mainly focused on the financial sector and limited our scope to Switzerland (i.e., context-bound). Second, our dataset covers a specific time (i.e., time-bound). Also, as with other qualitative research, the relationships uncovered between roles and organizational orchestration mechanisms, and the ability to govern frontier-level AI, raise questions about their generalizability beyond the specific context (i.e., method-bound). Thus, our insights are primarily applicable to organizations similar to ours. Future research should examine the applicability of these findings across different contexts to gain insights that might not be visible through this paper's setup.

6 References

Anthropic. (2025, October 15). *System Prompts*. https://docs.claude.com/en/release-notes/system-prompts

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing Artificial Intelligence. *MIS Quarterly*, *45*, 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Broek, E., Sergeeva, A., & Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*, 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Chen, Z., Ma, J., Zhang, X., Hao, N., Yan, A., Nourbakhsh, A., Yang, X., Mcauley, J., Petzold, L., & Wang, W. (2024). *A Survey on Large Language Models for Critical Societal Domains: Finance, Healthcare, and Law*.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550. JSTOR. https://doi.org/10.2307/258557

Emerson, R. M. (1962). Power-Dependence Relations. *Power in Modern Societies*. https://api.semanticscholar.org/CorpusID:53704572

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111–126. https://doi.org/10.1007/s12599-023-00834-7

Gioia, D., Corley, K., & Hamilton, A. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*, 15–31. https://doi.org/10.1177/1094428112452151

Horwitz, J. (2025, August 14). *Meta's AI rules have let bots hold 'sensual' chats with kids, offer false medical info*. https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, *37*(3), 101493. https://doi.org/10.1016/j.giq.2020.101493

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Yan, X., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2022). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, *55*. https://doi.org/10.1145/3571730

Landymore, F. (2023, December 21). *Car Dealership Disturbed When Its AI Is Caught Offering Chevys for $1 Each*. https://futurism.com/the-byte/car-dealership-ai

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*, 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Mäntymäki, M., Minkkinen, M., Birkstedt, T., & Viljanen, M. (2022). Defining organizational AI governance. *AI and Ethics*, *2*(4), 603–609. https://doi.org/10.1007/s43681-022-00143-x

Mayer, A.-S., van den Broek, E., & Karačić, T. (2025). 'Let Me Explain': A Comparative Field Study on How Experts Enact Authority Over Clients When Facing AI Decisions. *Journal of Management Studies*, *n/a*(n/a). https://doi.org/10.1111/joms.70022

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization*, *17*(1), 2–26.

Papagiannidis, E., Mikalef, P., & Conboy, K. (2025). Responsible artificial intelligence governance: A review and research framework. *The Journal of Strategic Information Systems*, *34*(2), 101885. https://doi.org/10.1016/j.jsis.2024.101885

Roy, A. (2024, January 26). *How GM's Cruise robotaxi tech failures led it to drag pedestrian 20 feet*. https://www.reuters.com/business/autos-transportation/how-gms-cruise-robotaxi-tech-failures-led-it-drag-pedestrian-20-feet-2024-01-26/

Salunke, S., Weerawardena, J., & McColl-Kennedy, J. R. (2011). Towards a model of dynamic capabilities in innovation-based competitive strategy: Insights from project-oriented service

firms. *Capabilities, Innovation and Competitive Advantage*, *40*(8), 1251–1263. https://doi.org/10.1016/j.indmarman.2011.10.009

Shneiderman, B. (2020). Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction*, 109–124. https://doi.org/10.17705/1thci.00131

Siggelkow, N. (2007). Persuasion With Case Studies. *Academy of Management Journal*, *50*, 20–24. https://doi.org/10.5465/AMJ.2007.24160882

Too, E. G., & Weaver, P. (2014). The management of project management: A conceptual framework for project governance. *International Journal of Project Management*, *32*(8), 1382–1394. https://doi.org/10.1016/j.ijproman.2013.07.006

Walsham, G. (1995). Interpretive case studies in IS research: Nature and method. *European Journal of Information Systems*, *4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9

Yildirim, N., Oh, C., Sayar, D., Brand, K., Challa, S., Turri, V., Walton, N., Wong, A., Forlizzi, J., McCann, J., & Zimmerman, J. (2023). *Creating Design Resources to Scaffold the Ideation of AI Concepts* (p. 2346). https://doi.org/10.1145/3563657.3596058