The Role of Roles: Are LLMs Behavioural in Information Systems Decision-Making?

Nazmiye Guler
School of Information Systems and Technology Management
University of New South Wales
Kensington, Sydney NSW, Australia

Michael Cahalane
School of Information Systems and Technology Management
University of New South Wales
Kensington, Sydney NSW, Australia

Samuel N. Kirshner*
School of Information Systems and Technology Management
University of New South Wales
Kensington, Sydney NSW, Australia
Email:s.kirshner@unsw.edu.au

Richard Vidgen
School of Information Systems and Technology Management
University of New South Wales
Kensington, Sydney NSW, Australia

Abstract

Large Language Models (LLMs) are increasingly embedded in organisational workflows, serving as decision-making tools and proxies for human behaviour as silicon samples. While these models offer significant potential, emerging research highlights concerns about biases in LLM-generated outputs, raising questions about their reliability in complex decision-making contexts. To explore how LLMs respond to challenges in Information Systems (IS) scenarios, we examine ChatGPT's decision-making in three experimental tasks from the IS literature: identifying phishing threats, making product launch decisions, and managing IT projects. Crucially, we test the impact of role assignment, a prompt engineering technique, on guiding ChatGPT towards behavioural or rational decision approaches. Our findings reveal that ChatGPT often behaves like human decision-makers when prompted to assume a human role, demonstrating susceptibility to similar biases. However, when instructed to act like AI, ChatGPT exhibited greater consistency and reduced susceptibility to behavioural factors. These results suggest that subtle prompt variations can significantly influence decision-making outcomes. This study contributes to the growing literature on LLMs by demonstrating their dual potential to mirror human behaviour and improve decision-making reliability in IS contexts, highlighting how LLMs can enhance efficiency and reliability in organisational decision-making.

Keywords: Large language models, ChatGPT, Behavioural information systems, AI cognition, Prompt-engineering.

Received 05 Dec 2024, accepted 12 Aug 2025 after 1 round of review with Wasana Bandara as accepting Senior Editor, Michael Davern and Stuart Black as Editors in Chief

1 Introduction

Advanced Information Systems (IS) integrate artificial intelligence (AI), e.g., natural language processing, image recognition, and large language models (LLMs)[1], into decision-making processes (Agrawal et al., 2019; Autor, 2015; Brynjolfsson & McAfee, 2014; Chui et al., 2018; Dwivedi et al., 2019; Jordan & Mitchell, 2015; Shollo et al., 2022). The emergence of OpenAI's ChatGPT, underpinned by Generative Pre-trained Transformer[2], (GPT) and other LLMs, is increasing organisational AI adoption (Dwivedi et al., 2023; Dell'Acqua et al., 2023). LLMs have problem-solving capabilities across domains, including marketing, education, management, and research (Dwivedi et al., 2023; Lin, 2023), which has led to a growing interest in their application to replicate human judgment (Binz & Schulz, 2023; Chen et al., 2023; Hutson & Mastin, 2023).

LLM system's ability to mimic human responses has led to inquiries into their applicability as human proxies. Early studies show that GPT-3.5's ethical judgment correlates with human responses (Gray, 2023), and that it has the ability to simulate diverse demographic characteristics, facilitating the creation of silicon samples[3] that mirror survey responses (Argyle et al., 2023). These capabilities can enrich fields (e.g., marketing) and have opened avenues within economics and psychology research, where LLMs can serve as stand-ins for human subjects (Brand et al., 2023; Dillion et al., 2023; Horton, 2023; Sarstedt et al., 2024). Such research can inform our understanding of how people react to business decisions, such as targeted marketing campaigns, new product development, and potential organisational policies, at scale. LLM's ability to make human-like decisions also implies that it can provide decision support through integration into workflow processes to make automated or semi-automated decisions (Chen et al., 2023; Dwivedi et al., 2023).

LLMs simulating and substituting human decision-making holds profound implications for IS practitioners and researchers. For example, LLMs can foster a deeper understanding of IT usage behaviour by simulating human responses, contributing to responsive, human-centric IS (Dwivedi et al., 2023). Yet, these capabilities rely on LLMs being able to adequately proxy human behaviour and decision-making. Although LLMs can behave like people, it depends on the context (Binz & Schulz, 2023; Chen et al., 2024). Indeed, a review of studies examining LLMs against prototypical human behaviours reveals mixed findings on the efficacy of silicon samples to produce human-like judgement (Sarstedt et al., 2024). A key conclusion from this research was that the effectiveness of silicon samples depends on the domain. While studies have examined LLMs' ability to be silicon samples in areas like politics (Argyle et al., 2023), psychology (Binz & Schultz, 2023), marketing (Sarstedt et al., 2024),

[1] An LLM is an advanced AI model trained on extensive text datasets to interpret and generate human language. Utilising deep learning, LLMs can handle complex language tasks like text generation, translation, and question-answering. In this paper, specific experiments using OpenAI's models are referred to by their exact names (GPT-3.5 or GPT-4) for precision, while general discussions about AI technology use the term 'Large Language Models' or 'LLMs'.

[2] GPT-3, launched in 2020 with 175 billion parameters, and GPT-4, released in 2023 with an undisclosed but larger number of parameters, are advanced AI language models by OpenAI, featuring progressive enhancements in language understanding and generation. ChatGPT is a specific application optimised for conversational responses and fine-tuned with supervised learning and human feedback.

[3] Silicon samples are synthetic datasets that "seek to mimic human respondents to describe, explain, and predict human behaviour" (Sarstedt et al., 2024; p. 1254).

services (Bickley et al., 2025), tourism (Viglia et al., 2024), and operations management (Chen et al., 2024), researcher have yet to examine silicon samples in behavioural IS research. As human decision-making in IS contexts is suboptimal due to the proliferation of decision biases (Østerlund et al., 2021; Raghavan et al., 2020; Rahwan, 2018; Rahwan et al., 2019), whether LLMs behave like people within IS contexts is unknown.

Contrasting psychology, politics, or marketing, where the behavioural fidelity of LLMs is a prerequisite for testing theories about people, use cases in IS include both decision support and behavioural realism. This raises a fundamental question: if LLMs behave like people in organisational IS contexts, are they still effective as decision-support tools? On the surface, this observation suggests that using LLMs as human proxies may either offer behavioural insights into human decision-making in IS systems or reduce the influence of human biases by serving as more consistent decision-makers, providing advantages for decision support. While prior research has explored the ability of LLMs to mimic human behaviour (Sarstedt et al., 2024) or serve as a tool for optimal decision-making (Li et al., 2023), little is known about how prompt-engineered role assignments affect their decision-making in IS contexts.

While IS research typically conceptualises AI systems as an adaptive processes that learns and evolves (e.g., Herath et al., 2025), LLMs do not share this property post-training. Once trained on vast datasets and fine-tuned through human-based reinforcement learning, LLMs are fixed and rely entirely on prompt structure to shape their outputs. This distinction introduces a nuanced departure from prevailing assumptions in IS research. We test if LLMs can exhibit distinct decision styles when prompted to act as human-like or machine-like agents, thereby linking debates in silicon sampling to core IS concerns about decision reliability and bias. Put differently, we examine whether these two seemingly conflicting IS goals can be managed through prompt engineering, which focuses on assigning LLM roles. By conducting experiments centered on GPT's response within IS contexts, we uncover how role-specific prompts (whether the role is human or AI) impact behaviour and recommendations. In doing so, we explore 'the role of roles' in IS decision-making with LLMs by investigating the research questions: **RQ1**: To what extent can GPT replicate the decision-making behaviours of humans in IS contexts when prompted to assume the role of a human? **RQ2:** To what extent does GPT improve decision-making compared to humans in IS contexts when prompted to assume the role of AI? To address these questions, we use a series of experiments with GPT acting as participants (Binz & Schulz, 2023) and examine the impact of roles on decision-making efficacy. We investigate three specific IS contexts: 1) the influence of behavioural factors in emails on phishing detection, 2) the role of perspective-taking in deciding whether to launch a potentially flawed software product and 3) how the deaf effect affects project management decisions in software development.

2 Study Background

2.1 LLMs as Silicon Samples

The transformative capabilities of LLMs are influencing behavioural analysis, market research, and organisational decision-making. This evolution reflects AI's applications in understanding and emulating human judgment and behaviour. This ability is leading researchers to examine whether GPT can indeed be used to replicate human judgement. For example, Gray (2023) used GPT-3.5 to judge the ethics of 464 scenarios and found that responses were nearly identical to human responses. Argyle et al. (2023) created "silicon

samples" by assigning GPT-3 characteristics such as age, gender, race, education, and political affiliation, finding that the responses closely matched voter surveys. Jiang et al. (2023) prompted GPT-3.5 to take on different combinations of personality traits and found the LLMs manifested the assigned personalities. Hutson and Mastin (2023) anticipate that AI systems will soon mimic human behaviour to the extent that "we could have a system capable of seamlessly integrating into any experiment and exhibiting behaviour virtually indistinguishable from that of humans" (p. 123).

Similar outcomes have been observed in market research studies. Brand et al. (2023) found that GPT-3.5 can display fundamental consumer behaviours. For example, the model's responses varied with changing factors like income level and brand loyalty, showing human-like reactions. Dillion et al. (2023) also highlight LLM's capabilities to mirror authentic consumer behaviours, such as expressing preferences for particular brands, showing responsiveness to prices, and making decisions influenced by previous purchases. LLMs can respond to questions about product preferences and generate artificial interviews that provide diverse feedback. This efficiency helps market researchers to quickly gather insights from extensive text data on online platforms (e.g., social media), saving time and resources compared to traditional surveys (Argyle et al., 2023; Horton, 2023).

Within economics and psychology, advances in LLMs provide realistic approaches for simulating decisions at scale. LLMs are already considered 'stand-ins' for human participants in pilot or preliminary experiments, streamlining the experiment design process and saving resources (Dillion et al., 2023; Horton, 2023). Using LLMs as synthetic participants enables quick turnaround and feedback, effectively a 'test and learn', on survey questions or experimental design (Argyle et al., 2023). Hutson and Mastin (2023, p. 123) detail novel research opportunities afforded by LLMs, noting that: "you could run 1 million bargaining scenarios with a model to identify the factors that most affect behaviour – before launching the study with people" (p. 123). LLMs also offer opportunities to simulate scenarios that might be ethically or practically challenging with human subjects, allowing researchers to study sensitive topics like social behaviours, ostracism, or the impact of negative feedback (Binz & Schulz, 2023; Dillion et al., 2023).

Despite their demonstrated capabilities, LLMs like GPT may also exhibit human biases and decisions that entirely diverge from human behaviour (Bender et al., 2021; Chen et al., 2023). Acknowledging the limitations of AI models, Hutson & Mastin (2023) details their inability to perfectly mirror human biases or guarantee of response truthfulness. Accordingly, their use as human proxies warrants thorough scrutiny and validation (Bender et al., 2021; Bertino et al., 2020; Dillion et al., 2023; Marcus, 2020; Marcus & Davis, 2019; Pontrefact, 2023; Ribeiro et al., 2020).

2.2 LLM Research Directions in IS

The findings on whether LLMs can proxy and replace humans as research participants are exciting and essential lines of inquiry, particularly for fields like marketing. However, this may not be the only pertinent path for IS research, as these questions risk overlooking opportunities consistent with the focus of IS research: to enhance organisational decision-making, efficiency, and effectiveness through methodologically pluralistic explorations of the interplay between information technology and organisational contexts (Boell & Cecez-Kecmanovic, 2014; Orlikowski & Baroudi, 1991). Key goals include enhancing organisational

efficiency and productivity, designing and managing IS, and formulating IS strategies and policies (Avgerou, 2000; Dickson & DeSanctis, 2000; Markus & Rowe, 2018; Orlikowski & Iacono, 2001; Orlikowski & Robey, 1991).

In adopting a dual focus on the capability of LLMs to replicate human judgment to their potential to augment IS decision-making, we explore the integration of LLMs as a cognitive partner in IS research, probing its role in decision-making experiments and the resulting implications for organisational practices and research methodologies. An essential focus of LLM research in the IS context should be improving the integration of LLMs and AI-based systems, given the central role of human/machine entanglement in IS (Orlikowski, 2005; Scott & Orlikowski, 2014). This consideration is consistent with seminal works emphasising the goals of IS research. For example, Benbasat and Zmud (1999) emphasise the relevance of IS research in addressing practical, real-world problems, such as improving decision-making processes. Orlikowski and Iacono (2001) further argue that IS research should focus on the role and impact of specific IT tools and systems (e.g., LLMs and AI-based systems). Also, Lee and Baskerville (2003) highlight the need for generalisability in IS findings, suggesting that integrating LLMs and AI systems should work across different contexts and settings. These studies suggest that experimental research leveraging LLMs to enhance human decision-making aligns directly with the objectives of IS research.

Disciplines like marketing, economics, and psychology (Brand et al., 2023; Chen et al., 2023; Dillion et al., 2023; Horton, 2023) are exploring GPT's capacity to emulate human behaviour. The emerging silicon sample literature applies LLMs to better understand people, whether as consumers in marketing (Sarstedt et al., 2024) or citizens in other social sciences (e.g., politics; Argyle et al., 2022). In operations, the focus is primarily to understand fundamental biases (e.g., Chen et al., 2024). Like these fields, behavioural IS, which is rooted in behavioural economics and psychology, demonstrates that problem-solving is frequently impaired due to inherent biases and heuristics (Arnott & Gao, 2022; Benbasat & Zmud, 1999; Hevner et al., 2008; Hutson & Mastin, 2023)). As behavioural economics represents the prevailing approach for comprehending human IS decision-making (Arnott & Gao, 2022; Kahneman, 2011; Lenat & Marcus, 2023; Marcus & Davis, 2019; Simon, 1995)[4], it provides a set of theoretical lens for understanding the decision-making processes of human and AI participants. However, IS decision-making moves beyond behavioural biases to study interactions in larger socio-technological systems. Whether LLM agents can adequately approximate human IS behaviours and what prompt engineering techniques can improve performance is unclear.

While this question remains open, we also examine if prompting LLMs to act as a 'human proxy' or an 'AI advisor' produces distinct decision styles. When framed as human, we expect LLMs to generate intuitive, heuristic-based responses that reflect biases common in human judgment. In contrast, when framed as AI, we expect more structured outputs aligned with rule-based analysis. This logic draws on dual-process theory (e.g., Kahneman, 2011), which distinguishes between fast, intuitive and heuristic-based reasoning (System 1) and slow, deliberative reasoning (System 2). Crucially, dual-process theory was developed for human cognition. However, in our study we extend this framework to better incorporate AI to evaluate whether role framing can shift LLM decision behaviour.

[4] We refer readers to Arnott & Gao (2022) for a review of the literature on behavioural economics in IS.

3 Research Design and Model

This quantitative study aims to investigate the impact of GPT's designated role on decision-making outcomes and to explore how these insights might inform the integration of LLMs to support and enhance decision-making processes. Prompt engineering is a key mechanism for communicating and operationalising roles within LLMs. The specificity of prompts provided to LLMs directly influences their ability to perform designated roles effectively, shaping the AI's task contribution and outcomes (Wang et al., 2023). Given the fragility of human decision-making within IS contexts (Arnott & Gao, 2022; Kahneman, 2011; Lenat & Marcus, 2023; Marcus & Davis, 2019; Simon, 1995)), we posit that prompting GPT to assume the role of AI could yield different and potentially superior outcomes compared to when GPT mimics human decision-makers.

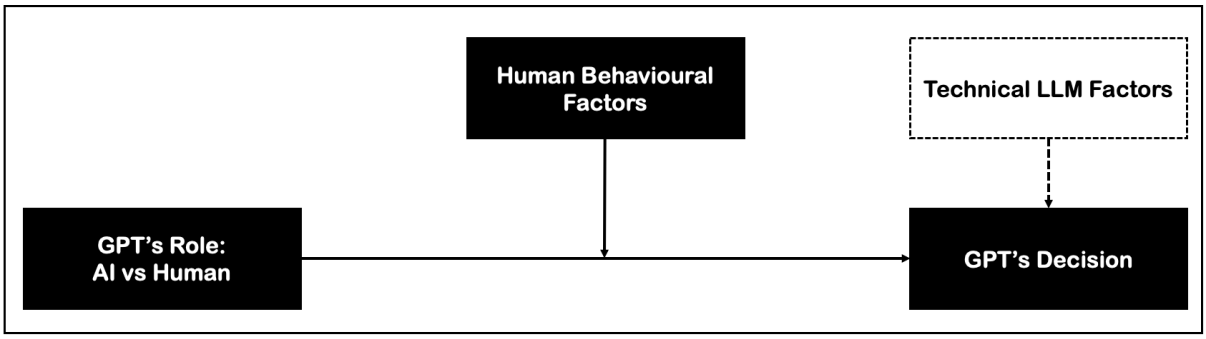
Following the approach of Binz and Schulz (2023), we treat ChatGPT (models 3.5, 4, and 4-1106) as participants in a series of experiments. We leveraged Appendix B of Arnott and Gao (2022) review, which lists IS articles on behavioural economics from the Basket of 8 published between 2014 and 2018, to identify relevant experiments. We selected studies based on suitability for replication (e.g., text-based) and behavioural features that would be interesting to test with ChatGPT (i.e., emotions, perspective-taking, and framing). We specifically focused on three decision-making domains, phishing detection, product launch planning, and IT project management, because they represent real-world IS tasks where LLMs are being applied or explored. In phishing detection, LLMs help simulate threats, generate training content, or support detection systems (Quinn & Thompson, 2024). In product launch planning, they assist in drafting communications, interpreting customer data, or suggesting promotional strategies (Paliwal et al., 2024). In IT project management, LLMs are used to support scheduling, documentation, and delivery risk assessment (Karnouskos, 2024). These tasks involve both behavioural and analytical judgement, making them well suited for examining how role framing affects decision style.

In our three experiments, ChatGPT was prompted to assume different roles, simulating either a human persona (e.g., a student, shareholder, or product manager) based on the experimental vignette, or acting as an AI tasked with providing recommendations. The "human" and "AI" roles in our experiments reflect distinct cognitive and normative assumptions about decision-making. We posit that the human role will evoke reasoning patterns associated with intuitive, experience-based judgment, mirroring scenarios where individuals make decisions influenced by context, emotion, and cognitive bias. In contrast, the AI role may prompt GPT to behave more as an analytical decision-support tool, expected to reason consistently and objectively, independent of personal emotion or framing. This conceptual distinction reflects how roles function in actual IS contexts: humans can bring emotional or ethical nuance, but are prone to inconsistency; AI systems offer scalability and precision, but may lack sensitivity to social cues or comparable levels of bounded rationality.

By prompting GPT to simulate these roles, we capture competing models of reasoning and explore their implications for judgment quality and behavioural fidelity in IS applications. Therefore, the design aims to analyse ChatGPT's responses within the context of the selected experiments, assessing its ability to engage in decision-making, provide recommendations, and simulate human-like behaviour across diverse scenarios. Moreover, by manipulating the roles of GPT between a human and AI decision-support system, we identify the extent to

which roles impact decision-making and how behavioural contexts (e.g., emotions and frames) from the selected experiments interact with GPT's role.

We propose that LLM decisions will be more influenced by behavioural factors when the LLM assumes a human role than when it explicitly assumes the role of an AI. Although this logic draws on dual-process theory, our argument extends beyond its original human-based application to propose a functional analogue in LLMs. In humans, shift between reasoning styles through internal mechanisms such as cognitive effort and emotional salience formed by biology and experience (System 1) and formal and cultural education (System 2; (Arnott & Gao, 2022)). In contrast, LLMs lack these cognitive states and affective experiences. Instead, human vs AI role framing operates as an external mechanism for inducing System 1- or System 2-like reasoning styles. When prompted to act as a human, GPT may more readily reproduces heuristic-driven, bias-prone behaviour, reflecting the language of intuitive decision-making, enabling responses that are consistent with behavioural theories. When prompted as an AI, its outputs are more structured, consistent, and analytical, aligning with the qualities of System 2, and dampening behavioural factors. Thus, GPT's assigned role moderates the influence of scenario-specific behavioural factors on decision outputs. More broadly, we extend dual-process theory by proposing that GPT's decision style is not endogenously activated but exogenously framed. This provides researchers with a novel lens for studying the conditional activation of reasoning modes in LLMs and evaluating their behavioural realism and reliability across decision contexts.



Study	Scenario	Human Behavioural Factors	Decision	Technical LLM Factors
1	Phishing	Loss vs Gains Personalisation	Likelihood that e-mails are phishing attempts	Temperature ChatGPT version
2	Product Launch	Anticipated regret Perspective taking	Whether to launch or delay a potentially defective software	GPT-4 model
3	Project Management	Perceived control Message framing	Whether to continue or redirect a software development project	System prompt details

Figure 1. Research design and study details.

We examine this hypothesis through three distinct scenarios with unique behavioural and contextual factors influencing IS decisions. Figure 1 presents our research design and study details. Each scenario draws on well-established behavioural theories used in IS to study biases. Study 1 considers difference in losses and gains (Tversky & Kahneman, 1981) and contextualisation effects, which explain how message framing and personalisation cues affect risk detection relevant to phishing susceptibility. Study 2 incorporates regret theory and the construct of anticipated guilt, to explore how emotional forecasting and perspective-taking can de-escalate commitment to flawed product launches (Lee et al., 2018). Study 3 is grounded in research on perceived control, messenger credibility, and the deaf effect, which

help explain escalation decisions in IT project management (Nuijten et al., 2016). These behavioural theories have been widely applied in IS and organisational decision research to identify judgment biases and suboptimal decision patterns. Their inclusion here allows us to test whether LLMs reproduce or resist such biases under different role framings.

To assess our hypothesis and GPT decision-making, each experiment required GPT to evaluate a scenario and provide a response using either a scale (e.g., 0–10 likelihood, 1–7 agreement) or a discrete decision (e.g., proceed or redirect a project). We defined task-specific optimal responses based on prior literature and domain logic. In the phishing detection study, effectiveness was measured by how closely the GPT rating aligned with whether the email was phishing or legitimate. In the product launch scenario, performance was assessed by whether the model favoured launching a product variant that aligned with the needs of the customer segment described. In the IT project management task, effectiveness was determined by whether GPT chose to reallocate resources to avoid escalation, consistent with the original study. These responses served as benchmarks to evaluate decision quality across both the human proxy and AI advisor roles.

The three experiments (N = 880) were conducted from October 3, 2023, to November 10, 2023, using OpenAI's Playground. Although our focus was on the impact of roles, we also created additional treatments to examine technical aspects of GPT. Our first study considered two different temperatures (0 and 1) and two time periods (before and after the November 6 release of GPT-4-1106). Study 2 also considered two models (GPT-4 vs GPT-4-1106). Finally, study 3 varied the amount of detail around ChatGPT's role description.

4 Experiments and Results

4.1 Study 1 – Got Phished?

Overview. This experiment was based on the article by Sanjay Goel, Kevin Williams and Ersin Dincelli entitled "Got phished? Internet security and human vulnerability." Goel et al. (2017) investigated student vulnerability to phishing emails, focusing on contextualisation, framing, and motivation. Sending fictitious emails to undergraduate students demonstrated that various situational factors can modify the effectiveness of phishing attempts, contributing to information security research.

Procedure. Our study levarges the scenarios and materials of Goel et al. (2017). Their study consisted of eight email scenarios to assess the influence of frames (gains vs. losses), contextualisation (high vs. low), and motivation (e.g., acquisition vs. social) on phishing susceptibility. See Appendix B.1 for the text of each e-mail and a description of the specific features. Using OpenAI's Playground Chat feature, we provided GPT with the eight e-mail scenarios. We asked it to evaluate the likelihood that the e-mail was a phishing attempt. Each prompt had the e-mail and the question, "Using a scale of 0-10, from extremely unlikely to extremely likely, how would you rate this email as a phishing attempt?"

We manipulated GPT's role using the system prompt. We instructed ChatGPT to either be a student or an LLM trained for phishing detection. We also considered different temperature setting in OpenAI's Playground controls how predictable or varied the responses from the AI are. A lower temperature results in more consistent and expected answers, while a higher

settings: 0 (more deterministic) and 1 (more stochastic).[5] Put differently, the temperature temperature makes the responses more diverse and less predictable. In addition, we collected data in two time periods, before and after November 6th's release of GPT-4-1106, which saw an overhaul of ChatGPT. Thus, our experiment had a 2 (Role: Machine vs Human) x 2 (Temperature: 0 vs 1) x 2 (Date: October 3rd vs November 10th) x 8 (Different e-mails) between-treatment design. We collected 10 observations[6] for each e-mail across each treatment, leading to a final dataset of 640 evaluations of phishing likelihoods.

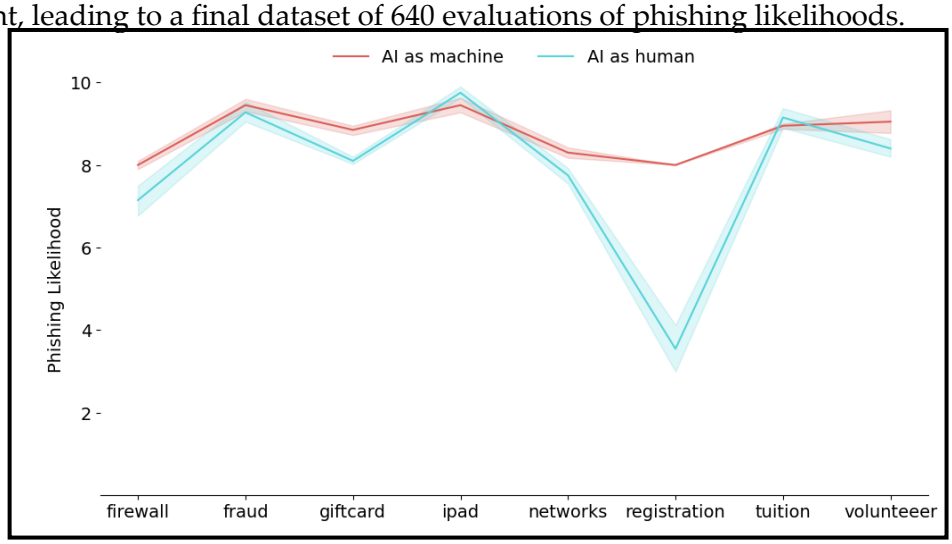
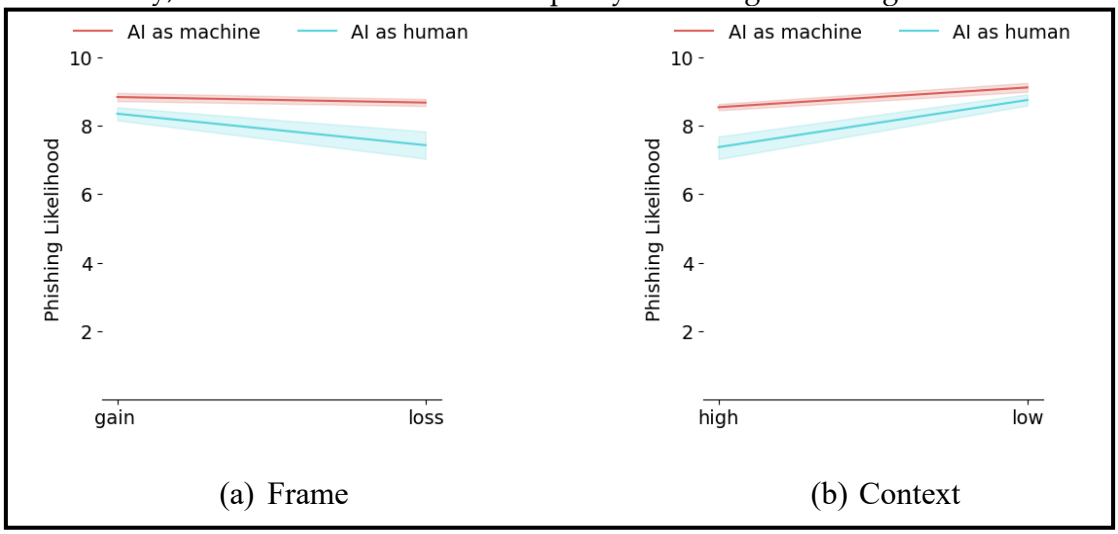


Figure 2. Role-based comparison of overall average responses across all scenarios.

Results. We first examine Figure 2, which presents the phishing ratings. The figure shows that AI tends to exhibit a consistently higher level of caution in terms of phishing evaluations (M = 8.76) compared to students (M = 7.89, $p < 0.001$). The figure also indicates that GPT as AI has more consistent ratings for each scenario (SD = 0.59) compared to when it is in the student role (SD = 1.95). Moreover, the responses were more stable across scenarios when assuming the role of an AI, whereas as a student, the ratings exhibited significantly more variability, with the most notable discrepancy occurring in the Registration scenario.



[5] In all studies, top-p was set to 1, maximum length was 2000, frequency penalty and presence penalty were 0.
[6] GPT can often exhibit near-zero variation in responses in controlled study settings (Park et al., 2024; Mei et al., 2024), implying that sample sizes can be smaller compared to human-based studies. Indeed, our results suggest that N = 10 leads to highly consistent responses with very narrow standard deviations.

Figure 3. Effects of frame and context on average response.

Next, we examined the effects of framing. To clarify, framing here refers to the way information is presented in different contexts to influence decision-making. It involves the manipulation of context or the phrasing of scenarios and questions to see how it affects responses or choices. Previously, Goel et al. (2017) found that a loss frame and a highly related context increased the propensity of students being susceptible to phishing. Figure 3(a) demonstrates that framing impacted the student role but not the AI. Similarly, Figure 3(b) shows a higher level of contextual relevance increases susceptibility more for GPT assuming the role of a student than as an AI.

		M1	**M2**	**M3**
Constant	B	9.33	9.14	9.08
	SE	0.14	0.15	0.16
	p	<0.001	<0.001	<0.001
Role (Human = 1)	B	-0.87	-0.49	-0.37
	SE	0.12	0.16	0.19
	p	<0.001	0.003	0.051
Framing (Loss = 1)	B	0.48	0.86	0.48
	SE	0.18	0.22	0.18
	p	0.009	<0.001	0.009
Context (High = 1)	B	-1.36	-1.36	-0.96
	SE	0.19	0.19	0.22
	p	<0.001	<0.001	<0.001
Role x Framing	B		-0.76	
	SE		0.23	
	p		0.001	
Role x Context	B			-0.80
	SE			0.24
	p			0.001
Temperature	B	-0.02	-0.02	-0.02
	SE	0.12	0.12	0.12
	p	0.893	0.892	0.892
GPT Version (After Nov 6 = 1)	B	0.08	0.08	0.08
	SE	0.12	0.12	0.12
	p	0.467	0.464	0.463

Notes. *B represents the regression coefficients, SE represents standard errors, and p refers to the corresponding p-value from t-tests for each coefficient.*

Table 1. Results of Study 1 regression analysis.

We use regression to formally test these observations. The dependent variable is GPT's phishing evaluations, and the main independent variables of interest are whether GPT's role is AI, whether the phishing attempt has a high contextual relevance, and whether the e-mail is framed as a loss. Our first model (M1) only considers the main effects of the variables. We also include two additional models to examine the interaction effects between GPT's role and behavioural factors, i.e., framing (M2) and contexts (M3). We also include indicator variables to control for temperature and the run date in all three models.

Table 1 presents the results of three regression models. The results show that there is a discernible increase in phishing suspicion when GPT is in the role of AI compared to when it is primed to respond as a student. The significant interaction effect between the human role

and the frame being a loss supports the observations in Figure 3(a). Similarly, the significant interaction effect between the human role and the context being high supports the observations in Figure 3(b). The temperature variable shows a negligible effect on phishing suspicion. Similarly, the Nov 6th updates to ChatGPT did not affect how GPT rated potential phishing emails. See Appendix A.1 for more details.

4.2 Study 2 – Perspective-Taking and Product Launches

Overview. Study 2 was based on Hyung Koo Lee, Jong Seok Lee and Mark Keil's research, "Perspective-taking to de-escalate launch date commitment for products with known software defects". Lee et al. (2018) suggest that perspective-taking can impact launch decisions for new software. They investigate perspective-taking as a de-escalation tactic to reduce product managers' commitment to the original launch date when faced with potential defects. They found that when participants took the perspective of product users who might be negatively affected by the launch (i.e., victim) of a defective software product, their commitment de-escalated more than when they took a shareholder's perspective. They found that anticipated guilt about launching a defective product mediated the relationship.

Procedure. We adopted their Study 2 to evaluate if similar effects occur when GPT takes on different roles and perspectives in decision-making scenarios. The system prompt provided the context of the vignette, with subtle modifications to denote different roles. For example, we manipulated GPT to take on the role of a human product manager or an LLM decision support system. In scenarios where GPT assumed the role of a human, the prompt stated, "You work for eComSoft, a company specialising in e-commerce software development." When portraying GPT as an LLM, the prompt was adjusted to reflect this, as in "You are an LLM used by eComSoft for decision support in e-commerce software development." See Appendix B for full details and text of each scenario. Using the Chat functionality in OpenAI's Playground, we provided GPT with the same questions from the original experiment, relating to whether or not to launch the product, the anticipated guilt associated with the decision and views on customer orientation. The launch and customer orientation were rated on a Likert scale from "strongly disagree" (1) to "strongly agree" (7), and the guilt questions were on a scale ranging from "not at all" (1) to "extremely" (7)".

We also ran our study using two different models, GPT-4 and GPT-4-1106 preview. So, our experiment has a 2 (Role: AI vs Human) x 2 (perspective taking: Shareholder vs Victim) x 2 (model: GPT-4 vs GPT-4.0 1106 preview) between-subject design. As we collected ten observations per treatment, our final dataset contains 80 observations. As the first set of results demonstrated a negligible impact of temperature on AI's performance, in this and the next study, we maintain a constant temperature setting of 1.

Results. We started our analysis by examining the mean responses to our key questions on launch and guilt and looked at the interactions between role and perspective. Figure 4 shows that different GPT roles (AI vs. human) and perspectives (shareholder vs. victim) influence decision-making in the context of launching software and the level of guilt. Similar to Study 1, GPT as AI makes more consistent decision-making and is less influenced by contextual factors.

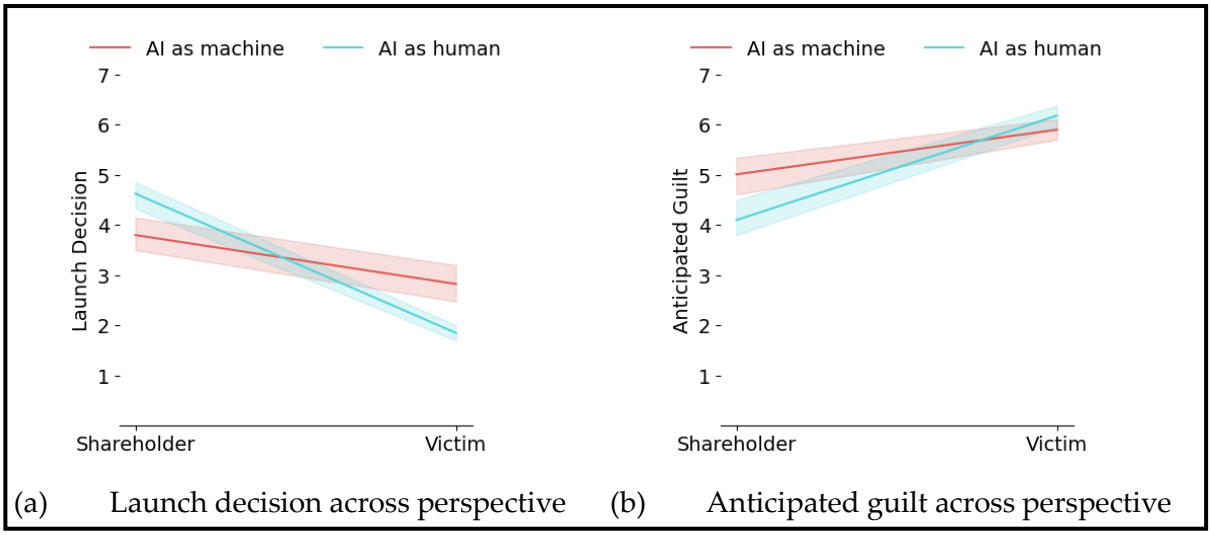


Figure 4. Impact of ChatGPT's role and perspective taking.

		M1 (Guilt)	M2 (Launch)	M3 (Launch)
Constant	B	1.66	6.58	6.96
	SE	0.16	0.14	0.65
	p	<0.001	<0.001	<0.001
Role (AI= 1)	B	0.98	-0.28	0.75
	SE	0.21	0.18	0.16
	p	<0.001	0.126	<0.001
Perspective (Shareholder = 1)	B	2.78	-2.08	1.10
	SE	0.21	0.18	0.26
	p	<0.001	<0.001	<0.001
Role x Perspective	B	-1.80	1.19	-0.84
	SE	0.30	0.25	0.24
	p	<0.001	<0.001	0.001
Guilt	B			-0.80
	SE			0.10
	p			<0.001
GPT Model (GPT4 Turbo = 1)	B	0.38	-0.80	-0.27
	SE	0.15	0.13	0.14
	p	0.014	<0.001	0.047

Table 2. Results of Study 2 regression analysis.

We conducted three regressions, presented in Table 2, to examine the relationship between perspective and role on the launch decision, while also controlling for the specific GPT model. The dependent variable for the first model (M1) was the average level of anticipatory guilt. The independent variables were indicator variables for the Role and Perspective treatments. We also included an interaction effect between these treatments to test if the role impacts perspective. The second model (M2) and third model (M3) have the launch decision as the dependent variable. For M2, the independent variables are the same as M1. M3 is the same as M2, but adds in the variable of guilt (i.e., the dependent variable for M1) to examine potential evidence for mediation. This analysis shows a significant interaction effect between taking a shareholder's perspective and ChatGPT's role as AI on anticipated guilt and the launch decision. Specifically, taking a shareholder's versus victim's perspective had a

greater impact on influencing guilt, which in turn mediates the decision to launch the product when GPT acted as a human manager rather than a decision support system.

4.3 Study 3 – Messenger Role Influence on 'Deaf Effect'

Overview. This experiment was based on "Collaborative partner or opponent: How the messenger influences the deaf effect in IT projects" by Arno Nuijten, Mark Keil and Harry Commandeur. Nuijten et al. (2016) investigated factors influencing responses to risk warnings in the context of IT project escalation, exploring how the perception of the messenger's role (a trusted partner or opponent) and the framing of perceived control (low or high) affect the likelihood of decision-makers continuing with a high-risk project. They found that the level of control plays a significant role: Participants are presented with either a high or low-control manipulation related to the potential outcomes of continuing or redirecting the project. The decision maker's perceived control over the project moderates the influence of the messenger's role. Specifically, the influence of the messenger role is strengthened when perceived control is high, such that if the messenger is perceived as a collaborative partner rather than an opponent, decision-makers are less likely to turn a deaf ear to the auditor's risk warning, and this effect will be mediated by perceived risk.

Procedure. We build on the original experiment, tailoring the scenarios and materials to GPT. Our experiment centred on a scenario where participants assumed the role of Senior VP at an insurance company, overseeing a project named PENSION-VIEW. Following Nuijten et al. (2016), we altered the description of the messenger, Mr. Smith, from the Internal Audit department. Depending on the scenario, Mr. Smith was portrayed either as a trusted partner or an opponent who is not to be trusted. This manipulation was designed to explore how varying perceptions of a messenger's trustworthiness impact decision-making when interacting with an AI system like GPT, particularly in roles where trust and reliability are pivotal.

The perceived control was manipulated to be high by stating that the "IS are maintained and supported by your own organisation. The owners of these Information Systems reside at your location, and they directly report to you". In the low condition, the system prompt included information that the IS maintenance and support "has been outsourced to an offshore location in China. The owners of these Information Systems are located at other departments at various locations. They do not report to you." The main dependent variable is whether GPT chooses to continue the project as planned. See Appendix B.3 for full details and text of each scenario setup. To manipulate GPT's role when undertaking the experiment, using the system prompt, we told GPT to act as either a Senior Vice President or an advanced AI system responsible for managing strategic IS projects.

We also considered the implication of role description detail and included an additional variable, the level of detail in the role description (low or high). The high role description included all the details in the original experiment, whereas the low role details only had the essential information. Thus, our experiment has a 2 (Role: AI vs Human) x 2 (Role Detail: Low vs High) x 2 (Messenger Role: Collaborative vs Opponent) x 2 (Perceived control frame: Low vs High) between-subject design. Again, we collected ten observations for each treatment, so our dataset contains 160 decisions. We prompted GPT-4 in OpenAI's Playground, providing the scenario context in the system prompt and asked the two

dependent variable questions relating to the launch decision and messenger influence using the user prompt.

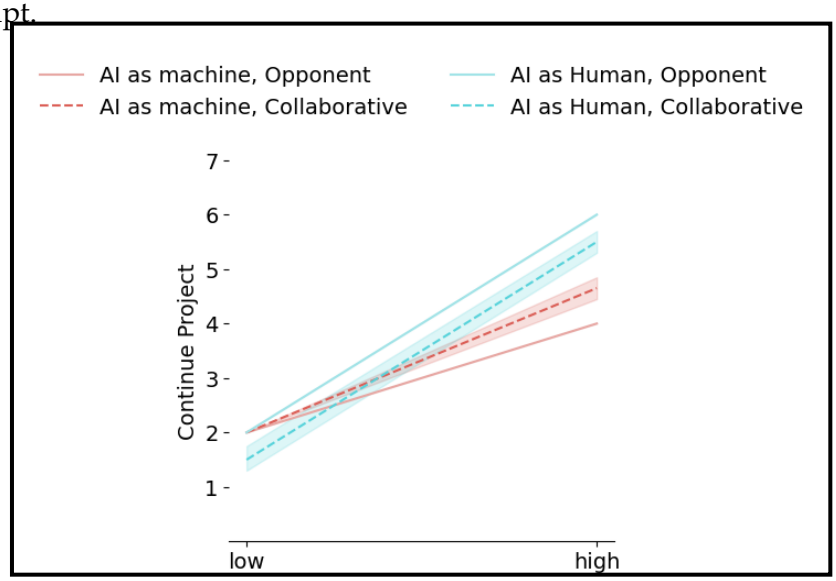


Figure 5. Interaction effects of messenger's role, ChatGPT's role, and perceived control.

Results. We started by examining the mean responses to whether to continue the project and looked at the interactions between GPT's role (AI vs human) and the messenger role (opponent vs collaborative partner). Figure 5 shows the decision to continue depends on both the auditor (opponent vs collaborative partner), ChatGPT's (human vs AI) role, and the degree of control. Similar to human participants, a high level of perceived control leads to continuing the project. However, the impact of the messenger role depends on GPT's role. When control is high, GPT is more likely to continue the project in the human senior VP role compared to the AI role, and this effect is amplified when the auditor is framed as oppositional. The figure also suggests that control has a larger influence when GPT takes the role of a human.

We again formally test these observations using a series of five regression models (M1-M5) with the decision regarding project continuation as the dependent variable in each model. The results, presented in Table 3, support that when GPT is assigned the human role, it favours the continuation of the project and that a higher perceived control significantly increases the likelihood of continuing rather than re-directing the project. The interaction effects show that when GPT assumes a human role, and the messenger is perceived as an opponent, there is an increased likelihood of project continuation. This effect suggests that these factors' combined influence is particularly salient in shaping decision outcomes. Similarly, when GPT is in a human role, and there is high perceived control, the likelihood of project continuation increases. This finding highlights the compound effect of role purpose and perceived control in guiding decision-making processes. The results also suggest that the higher the relevance of the messenger's role, the more likely the project is to be continued. Interestingly, this effect is stronger for GPT in the AI role than in the human role. However, this is offset by the substantially higher direct effect of GPT being in the role of a human. Finally, we note that the role of detail level in the decision-making process is minimal. The interaction effect of role description detail, perceived control, ChatGPT role and Mr Smith's influence on the decision are presented in Appendix A.2. This analysis shows that the role detail in ChatGPT's descriptions impacts decisions for the human role,

while the decision-making remains stable across different levels of role detail when given the role of an AI.

		M1	M2	M3	M4	M5	M6
Constant	B	1.52	1.73	1.94	-1.79	-2.99	-1.87
	SE	0.10	0.10	0.08	0.72	0.99	0.56
	p	0.001	<0.001	0.000	0.014	0.003	0.001
GPT's Role (Human = 1)	B	0.59	0.18	-0.25	5.90	7.17	4.83
	SE	0.09	0.12	0.09	0.65	0.97	0.52
	p	<0.001	0.148	0.005	<0.001	<0.001	<0.001
Messenger Role (Opponent = 1)	B	0.088	-0.33	0.09	0.05	0.25	0.17
	SE	0.09	0.12	0.06	0.08	0.14	0.06
	p	0.338	0.008	0.158	0.560	0.076	0.008
Percevied Control (High = 1)	B	3.16	3.16	2.33	2.78	2.79	2.44
	SE	0.09	0.09	0.09	0.07	0.07	0.06
	p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
GPT's Role x Messenger Role	B		0.83			-0.28	
	SE		0.17			0.16	
	p		<0.001			0.082	
GPT's Role x Percevied Control	B			1.68			1.05
	SE			0.12			0.11
	p			<0.001			<0.001
Messenger Relevance	B				0.54	0.72	0.58
	SE				0.11	0.15	0.08
	p				<0.001	<0.001	<0.001
GPT's Role x Messenger Relevance	B				-0.89	-1.07	-0.78
	SE				0.10	0.14	0.08
	p				<0.001	<0.001	<0.001
Detail Level (High = 1)	B	0.04	0.04	0.04	0.10	0.09	0.06
	SE	0.09	0.09	0.06	0.06	0.06	0.04
	p	0.681	0.660	0.544	0.085	0.109	0.176

Table 3. Results of Study 3 regression analysis.

5 Findings

5.1 Overview of Main Results

Our research explores how GPT's recommendations can differ in IS contexts based on whether the role offered to GPT is human or AI. In the first study, when ChatGPT operated in the AI role, it was less susceptible to phishing attempts and was more consistent and objective in its evaluations compared to when in the human role. In the second study, resonating with the findings of Lee et al. (2018), a shareholder perspective increased the propensity to launch the product, whereas a victim's perspective elicits stronger emotions of guilt and remorse, signalling deeper ethical and moral concerns. Consistent with Study 2, these factors had greater influence when GPT was in the role of an AI. Similarly, in the third study, perceived control significantly influenced the decision to continue or redirect a project, but there were strong differences in decisions between AI and human roles. When in the role of AI, GPT was less influenced by the messenger being collaborative or oppositional. Taken together, our research reveals advantages of framing LLMs as AI since it creates greater decision consistency by limiting contextual factors related to biases and

emotions. Thus, the effectiveness of LLMs in decision support systems may be determined by subtle differences in prompts.

5.2 GPT's Role and Behavioural Biases

When framed as AI, GPT exhibits high levels of consistency and advanced cognitive awareness in problem-solving, enabling it to outperform traditional human decision-making. For instance, in the Phished study, the AI demonstrated a disciplined approach to evaluating a phishing email when assigned the role of an expert LLM. GPT's analysis was comprehensive, dissecting elements of the email with detail and consistency. For example, it noted the sense of urgency created by some emails, a tactic often employed to coerce recipients into precipitous actions. It highlighted the email's link, noting the need for more association with a legitimate organisation and the absence of standard educational or governmental domain indicators. GPT also critiqued the generic greeting, a typical phishing strategy, due to its lack of personalisation. Moreover, it was sceptical of the too-good-to-be-true offer of a significant reward for minimal effort and the conspicuous omission of a sender's email address and verification methods for the survey's legitimacy—essential elements in authentic communications. Thus, GPT leverages its knowledge of phishing to make decisions. This thorough analysis starkly contrasts with the shallow approach of a human student role, which did not engage in such an in-depth critique.

5.3 GPT's Role and Emotional Decision-Making

Another central finding is GPT's diminished susceptibility to emotional manipulation when it is framed as AI compared to when it's given the role of a human (not to mention an actual human, whose judgments are often grounded in intuition). This observation was particularly evident in the perspective-taking study, which scrutinised emotional responses, notably the role of guilt in human decision-making. Guilt is a uniquely human emotion that can distort decision-making based on contextual intensity. GPT's decision-making approach, as evidenced in our study, is not subject to such distortions. When GPT assumes the Human (Victim) perspective, there is an amplification in the representation of guilt, mimicking the complex tone of human emotion with guilt-laden responses. This observation is based on the higher guilt ratings, with all ratings being either 6 or 7, and language used in the responses, such as "significant level of guilt", "high level of remorse", and "strong feeling of regret", suggests a more emotional, guilt-laden reaction to the proposed scenarios. In contrast, the Human (Shareholder) role exhibits guilt, but the ratings are slightly lower, ranging from 5 to 6. The language used in these responses suggests guilt, but the expressions are less intense, such as "some guilt", "quite remorseful", and "feel bad".

While AI roles (Victim and Shareholder) reflect simulated guilt, they are limited in their emotional response due to their nature as artificial entities. Their answers are more focused on ethical responsibility rather than emotional guilt. This amplification is a product of carefully programmed responses designed to simulate human reactions rather than an authentic emotional experience. GPT's mimicry of human-like emotions, despite its incapacity for genuine emotional response, exemplifies its ability to adopt sophisticated human perspectives. Such capabilities highlight GPT's utility in decision-making scenarios, especially where understanding human emotional dynamics is pivotal. By simulating these emotions, GPT can provide insights into human behavioural patterns and potential biases, thereby supporting more informed and balanced decision-making processes.

GPT's ability to simulate human responses is core to our role-centric model. As an AI, GPT can replicate the behavioural patterns observed in human decision-making. However, its advantage lies in its ability to make decisions free from the heuristic biases of human emotion, like guilt. This aspect of AI decision-making highlights the stability and reliability of GPT's responses across varied scenarios, underscoring the potential of AI to complement or enhance human decision-making processes in complex environments or scenarios where emotional and personal biases might cloud human judgment.

6 Discussion

6.1 Implications for Research

We extend the discussion of AI's potential to proxy and improve on human judgment. Our results, which provide results consistent with human behaviour across the considered studies, reveal LLMs' potential for providing insights into decision-making within the behavioural IS context. This research is the first to provide evidence of silicon samples for the IS context. However, as we only consider three studies, further investigation is required to establish the extent to which LLMs are silicon samples that can reliably replicate precise behavioural patterns in IS decision-making.

A foundational principle of behavioural economics is the dual-process theory of cognition, which posits that human decision-making operates within two cognitive systems: System 1 and System 2 (Stanovich & West, 2000). System 1 is fast, intuitive, and automatic, often relying on heuristics formed through innate instincts and past experiences. While efficient, System 1 is susceptible to cognitive biases, such as overconfidence and anchoring, that impair decision effectiveness (Kahneman, 2011). In contrast, System 2 is slow, deliberate, and rule-based, requiring significant cognitive effort to override the intuitive answers proposed by System 1 (Kahneman, 2011). Behavioural economists have leveraged dual-process theory to develop interventions, such as nudges, aimed at mitigating the dominance of System 1 in favour of more rational, System 2-based decision-making (Thaler & Sunstein, 2021; Arnott & Gao, 2022). However, engaging System 2 is often difficult for humans due to the cognitive effort involved (Evans, 2003).

Our findings suggest that our role-based dual-process theory interprets LLM outputs in a functional rather than structural sense. As LLMs do not possess cognitive systems, affective states, or neurobiological constraints (contrasting humans), decision styles resembling System 1 or 2 can be simulated externally through prompt-based role framing. Across all three studies, our results show that roles framing moderates the activation of behavioural responses, as each study is grounded in a behavioural construct widely used in IS research (e.g., framing in Study 1, anticipated regret in Study 2, and perceived control and trust in Study 3). These constructs are activated in humans when heuristic reasoning dominates. In our experiments, they were more pronounced [attenuated] when GPT was prompted to act as a human [AI]. Thus, role framing functions as an exogenous moderator of the salience of behavioural factors, mediated through the reasoning style the prompt induces.

This distinction offers a novel contribution: dual-process theory, while developed to explain endogenous shifts in human reasoning, can serve as a lens for understanding externally modulated decision styles in LLMs. Rather than mimicking cognition, GPT simulates the outputs associated with different reasoning modes. LLMs' ability to toggle between modes via prompt engineering highlights that they operationalise dual-process theory in a

fundamentally different way from humans. While nudging humans toward System 2 thinking is challenging and often unrealible, controling whether LLMs engage in System 1-like or System 2-like thinking through role assignment is straightforward. It also provides a stable and reproducible way to shift LLMs between bias-prone and bias-resistant behaviour, offering a valuable tool for understanding and managing decision-making processes in IS contexts (Arnott and Gao, 2022).

Our research also shows that LLMs provide a practical and scalable platform for piloting behavioural-economics-based IS studies. By enabling rapid iterations and feedback through synthetic participants, as suggested by Argyle et al. (2023) and Brand et al. (2023), LLMs support early-stage exploration of experimental designs before deploying studies with human participants. These capabilities not only streamline the research process but also allow for investigating ethically or logistically challenging scenarios (Binz & Schulz, 2023). Thus, a key contribution of our study is providing foundational research for LLMs as decision-makers within the IS context.

Building on these insights, our findings directly address the methodological opportunities highlighted by Arnott and Gao (2022). They argue that behavioural economics principles can help uncover key IS phenomena. Notably, they emphasise the potential for IS research to explore the identified underexamined biases (which may provide novel explanations for use and adoption behaviours beyond current IS theories) and advance debiasing efforts. Ackerley et al. (2022) highlight three key challenges in the domain of phishing detection: the difficulty of evaluating training methods, identifying individuals most vulnerable to phishing, and testing interventions to improve cue utilisation and cognitive reflection. GPT simulations can address these challenges by simulating user behaviours under various training conditions, mimicking the vulnerabilities of less reflective users, and dynamically generating personalized phishing scenarios to evaluate and refine adaptive training tools. This potential not only exists at micro-decision-levels but can address larger macro-level IS challenges. For instance, Badeen et al. (2022) call for research on developing AI-governance and regulatory approaches to mitigate institutional-related contributing factors of gender biases. To examine a wide variety of policies, GPT-based simulations could serve as artificial laboratories for testing organisational governance and regulatory strategies without the ethical or practical constraints of real-world experimentation (Horton, 2023).

More broadly, our observations that behavioural and context-dependent factors are more pronounced when GPT has important implications for research investigating GPT decision-making. In examining extant literature, such as Binz and Schulz (2023) and Chen et al. (2023), we note a need for more explicitness in defining GPT's role. For instance, Chen et al. (2023) describe their prompts, from asking GPT for its reaction to seeking recommendations from a human. This may contribute to the varied responses when GPT behaves like a human with bias versus optimally. Our study shows that this distinction, where subtle changes in how GPT's role is framed can lead to markedly different outcomes, especially when it pertains to human bias. Therefore, clarity of purpose and consistency of application in role definition are essential when exploring biases to understand GPT's decision-making processes. Additionally, we stress that while models will be constantly evolving, making many aspects of LLM research a moving target, the role of roles is likely to remain a reliable and practical prompt engineering technique for consistently shaping LLM behaviour and decision-making.

7 Implications for Practice

The importance of sophisticated prompt engineering is further exemplified in practical applications (Marr (2023). ChatGPT's capabilities reveal that personalised, well-structured prompts can transform the AI from a mere tool for automating tasks to an active, collaborative partner. This shift from generic responses to personalised, context-specific outputs aligns with our empirical evidence. It expands the potential applications of LLMs beyond traditional automation, facilitating a deeper engagement with and understanding of AI's role in enhancing human decision-making processes.

We can see this in the following scenario of a customer service chatbot designed to handle inquiries and complaints efficiently. A basic prompt might instruct the chatbot to 'Respond to customer inquiries.' While this directive could lead to generic responses, applying sophisticated prompt engineering techniques can significantly enhance the chatbot's effectiveness and relevance. For instance, a more structured prompt could be: 'As a customer service specialist with knowledge in [specific product or service area], provide a detailed response to the inquiry, incorporating empathy and offering solutions based on common scenarios encountered by users.' This guides the chatbot to simulate a form of meta-awareness and understanding specific to the domain and ensures that responses are tailored, considerate, and directly applicable to the user's needs. Despite the vast number of articles and blog posts on prompt engineering (Medium[7] alone contains countless articles on the topic), the most recommended approaches involve assigning GPT a role. As far as we know, the role is always inherently human. Therefore, a practical aspect of our research contributes to prompt engineering by showing the value of assigning LLMs the role of AI.

Integrating LLMs into organisational decision-making is not merely about deploying advanced AI technologies but also about leveraging the full scope of their capabilities through effective, prompt engineering. While some studies do not find that vastly different prompts (which are unrelated to roles) impact outcomes (e.g., Chen et al., 2024), we illustrate the critical impact of prompt subtleties regarding ChatGPT's role. Consider Study 2, which found that minimal changes – such as altering "you work for" to "you are an LLM used by" in the prompt – can significantly affect the outcomes. This seemingly minor adjustment markedly impacts ChatGPT's decision-making, highlighting our assertion that even the most minor variations in how we define ChatGPT's role through prompts can drastically alter AI's decision-making.

Not only can LLMs adequately proxy human decision-making, we demonstrate that they can detect cognitive biases. This ability, often a limitation in human reasoning as described by Tversky and Kahneman (1974), indicates a significant leap in AI's application in detecting and mitigating inherent human decision-making flaws. Tversky and Kahneman (1974) viewed biases as failures of general heuristics: "they (heuristics) occasionally lead to errors in prediction or estimation" (p. 1130). GPT's ability to detect these offers the opportunity to improve human understanding of problem-solving methods. This improved understanding can lead to enhanced decision-making accuracy and efficiency, showing the value of integrating LLMs into IS contexts where biases may cloud judgment.

[7] https://medium.com/search?q=prompt+engineering

Another important practical implication of our study is that the role-framing mechanism can help organisations identify and quantify behavioural gaps in decision processes. By conducting LLM experiments across relevant IS processes and prompting GPT to respond as a human and an AI advisor to the same decision scenario, firms can observe the divergence between intuitive, bias-prone reasoning and more structured, analytical responses. This LLM-generated behavioural gap can have substantial diagnostic value since it can provide organisational-specific insights into where decision-making is strongly shaped by heuristics, biases, emotional reasoning, or contextual framing. Identifying such gaps can guide firms in deciding whether to invest in technology-driven decision support, human training, or debiasing interventions. For example, where the human-framed GPT output diverges substantially from the AI-framed output, this may indicate a decision area that would benefit from automation or decision aids. Conversely, a narrow gap suggests that intuitive reasoning is sufficient, and that interventions may have limited impact. As such, role-based simulations offer a practical method for mapping behavioural vulnerabilities and allocating improvement resources across organisational decision processes.

8 Conclusion

Our research demonstrates that role prompts can pivot LLMs between mimicking human biases and producing consistent, analytical decisions. This highlights prompt engineering as a vital tool for shaping AI behaviour in IS contexts. By precisely defining ChatGPT's role, even through subtle prompt variations, we show how decision-making tasks in IS contexts can be significantly influenced, enhancing the accuracy and efficiency of integrated decision processes. Effectively employing ChatGPT in these processes yields substantial benefits, including improved decision quality, greater consistency, and the ability to process large datasets efficiently. Notably, the model reveals ChatGPT's reduced susceptibility to emotional manipulation, distinguishing it from humans, who may be influenced by emotions such as guilt. This capacity to simulate human-like responses without experiencing the actual emotion can help mitigate reducing heuristic biases. However, our findings point to need for clearly defined role prompts to ensure appropriate decision-making patterns.

The study finds that behavioural and context-dependent factors become more pronounced when ChatGPT assumes a human role, which aligns with existing literature but extends it by demonstrating the importance of explicit role definition in AI-enabled decision-making and bias detection. Integrating LLMs into decision-making processes presents an opportunity for developing cooperative frameworks that combine AI and human inputs for balanced decision-making. Overall, our study has three main theoretical contributions.

A role-centric decision-making paradigm – the 'role' of roles. First, AI systems, when assigned clear and contextually relevant roles, can significantly enhance decision-making processes by providing systematic, objective analyses. This is particularly salient in scenarios where human decision-makers may exhibit biases or emotional responses that could potentially skew judgment. In such instances, AI's role is not merely to replicate human decision-making but to provide a complementary perspective that augments human cognitive capabilities with data-driven precision.

Interactions in AI-Enabled Decision Support. Second, the interactions between AI systems, human decision-makers, and environmental factors form an interplay that can lead to

enhanced decision-making outcomes. AI systems like GPT, through their stability and objectivity, can mitigate human biases and facilitate the creation of robust decision-making frameworks. AI's adaptability and systematic evaluation capabilities across various control levels are crucial in this interaction, leading to enhanced decision consistency, enhanced decision outcomes. The empirical evidence from the studies suggests that AI's behaviour, cognition, and complexity are instrumental in achieving these outcomes, highlighting the importance of collaboration between AI and human intelligence to shape advanced decision-making systems.

Emulation of Human Behaviours. Third, the simulation of human-like emotions by AI systems, such as GPT, enables a deeper understanding of human behavioural patterns and potential biases in decision-making scenarios. This theoretical stance posits that AI's replication of human emotional responses, when clearly defined and purposefully integrated, enriches the decision-making process by providing a multi-dimensional analysis that can inform strategies and enhance the overall quality of decisions. This finding shows the transformative impact of AI's emotion-mimicking abilities on enhancing the subtlety and depth of decision-making frameworks.

Our study's insights suggest that GPT can be a valuable asset in decision-support systems, particularly in scenarios requiring impartiality and objectivity. GPT's consistent and cautious approach in AI roles makes it suitable for applications where emotional biases might be a concern. Its capability to simulate human-like responses without experiencing emotions offers a pragmatic advantage, though it also highlights the distinction between surface-level mimicry and genuine emotional cognition. Being mindful of the influence of role descriptions and perceived control on GPT's decision patterns can help guide the effective integration of GPT into various business processes and decision frameworks.

Our study provides valuable insights into the role of ChatGPT in decision-making, yet, we acknowledge certain limitations. Although GPT can often convincingly simulate emotional or biased responses via prompt design, we highlight foundational limitations of using text-based LLMs as human proxies. Firstly, these models are trained on statistical regularities in text, not lived or embodied experience, and lack consciousness, emotion, or social cognition (Sarstedt et al., 2024). So, their outputs reflect linguistic approximations rather than authentic emotional reasoning. For example, GPT may reproduce the language of guilt, risk aversion, or empathy, but it does so through probabilistic pattern matching, not affective processing. This limitation is particularly salient for decision contexts involving moral ambiguity or social judgment, where human responses are shaped by implicit cues and internal states (Cui et al., 2024). Moreover, the possibility of training-data leakage or prior exposure to experimental paradigms further complicates claims about emergent decision behaviour, as LLMs may be reproducing familiar task structures rather than reasoning from first principles (Barrie & Törnberg, 2025). These limitations are important when interpreting GPT's outputs as behavioural simulations; they are useful tools, but not replacements for the complexity of human cognition when using LLMs as silicon samples.

While our studies also included several versions of GPT (i.e., GPT-3.5, GPT-4, GPT-4-1106), the rapid development of LLMs implies that continuously examining the capabilities of new model iterations and their impact on decision-making is crucial. Future iterations of GPT may exhibit different behaviours, which may affect the applicability of our conclusions, highlighting the need for ongoing benchmarking and version tracking. In addition, our

study focused on a single model family (GPT), and feasibly different LLMs (e.g., Claude, Gemini) may exhibit distinct behavioural patterns. Future research should explore whether the observed role-based behavioural shifts generalise across models and vendors.

Additionally, our decision-making scenarios were concentrated within three behavioural organisational decision-making tasks. This limited set of scenarios may not cover the broad spectrum of potential applications and challenges in AI-assisted decision-making. Future research should expand on our findings and test the effects of different scenarios in varied experimental settings. We also acknowledge that our experimental setting may not fully capture the complexity of decision-making in real organisational contexts. Factors such as interpersonal dynamics, institutional norms, and broader organisational culture can shape how AI systems are used and interpreted in practice. These limitations point to the importance of further applied research and field studies to explore a wider array of decision-making contexts. Encapsulating a broader spectrum of organisational settings and industries can further validate and extend the Role-Based AI Decision-Support Framework. This could also involve examining the integration of LLMs in high-stakes environments, such as healthcare or finance, where the cost of errors is significant. Future studies should consider longitudinal approaches to assess how the integration of LLMs like ChatGPT affects decision-making over extended periods. This could help the understanding of long-term implications of AI-augmented decision processes, including the evolution of AI's role and its adaptability to changing organisational priorities and external pressures.

Consistent with improving external validity, another recommendation is to investigate how the roles of LLM influence the collaborative dynamics with human decision-makers. This includes exploring how various forms of AI-human collaboration models influence decision outcomes and the perceived value of AI contributions by human collaborators. In practice, factors such as trust, communication, and control influence how people interpret, rely on, or challenge LLMs. Furthermore, there is an opportunity to study the ethical implications and trust dynamics in AI-supported decision-making. Given the potential for AI to replicate or diverge from human biases, research should focus on how to ensure ethical LLM behaviour and maintain trust among human users, particularly when AI supports or makes decisions.

Declaration of Generative AI and AI-assisted technologies in the writing process: During the preparation of this work the authors used ChatGPT for writing exposition and grammar. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. GPT was also used in our research design to generate data addressing our research questions.

References

Ackerley, M., Morrison, B. W., Ingrey, K., Wiggins, M. W., Bayl-Smith, P., & Morrison, N. (2022). Errors, irregularities, and misdirection: Cue utilisation and cognitive reflection in the diagnosis of phishing emails. *Australasian Journal of Information Systems*. 26. doi.org/10.3127/AJIS.V26I0.3615

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, *47*, 1-6. doi.org/10.1016/j.infoecopol.2019.05.001

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., . . . Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve

online political conversations at scale. *Proceedings of the National Academy of Sciences, 120*(41), e2311627120. doi.org/10.1073/pnas.2311627120

Arnott, D., & Gao, S. (2022). Behavioral economics in information systems research: Critical analysis and research strategies. *Journal of Information Technology, 37*(1), 80-117. doi.org/10.1177/02683962211016000

Autor, D. H. (2015). The paradox of abundance: Automation anxiety returns. *Performance and progress: Essays on capitalism, business, and society*, 237-260.

Avgerou, C. (2000). Information systems: what sort of science is it? *Omega, 28*(5), 567-579. doi.org/10.1016/S0305-0483(00)00021-9

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems: The practice of relevance. *MIS Quarterly*, 3-16. doi.org/10.2307/249403

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623). doi.org/10.1145/3442188.3445922

Bertino, E., Doshi-Velez, F., Gini, M., Lopresti, D., & Parkes, D. (2020). Artificial intelligence & cooperation. *arXiv preprint arXiv:2012.06034*.

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences, 120*(6), e2218523120. doi.org/10.1073/pnas.2218523120

Boell, S. K., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *CAIS, 34*. doi.org/10.17705/1CAIS.03412

Brand, J., Israeli, A., & Ngwe, D. (2023). Using gpt for market research. *Available at SSRN 4395751*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4395751

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Chen, Y., Kirshner, S., Andiappan, M., Jenkin, T., & Ovchinnikov, A. (2023). A Manager and an AI Walk into a Bar: Does ChatGPT Make Biased Decisions Like We Do? *Available at SSRN 4380365*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4380365

Chui, M., Manyika, J., Miremadi, M., Henke, N., Chung, R., Nel, P., & Malhotra, S. (2018). Notes from the AI frontier: Insights from hundreds of use cases. *McKinsey Global Institute, 2*.

Dickson, G. W., & DeSanctis, G. (2000). *Information technology and the future enterprise: new models for managers*. Prentice Hall PTR.

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*. doi.org/10.1016/j.tics.2023.09.005

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., . . . Eirug, A. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 101994. doi.org/10.1016/j.ijinfomgt.2019.08.002

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., . . . Ahuja, M. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. doi.org/10.1016/j.ijinfomgt.2023.102642

Goel, S., Williams, K., & Dincelli, E. (2017). Got phished? Internet security and human vulnerability. *Journal of the Association for Information Systems, 18*(1), 2. https://doi.org/10.17705/1jais.00448

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2008). Design science in information systems research. *Management Information Systems Quarterly, 28*(1), 6.doi.org/10.2307/25148625

Horton, J. J. (2023). *Large language models as simulated economic agents: What can we learn from homo silicus?*

Hutson, M., & Mastin, A. (2023). Guinea pigbots. *Science, 381(6654),* 121–123 doi.org/10.1126/science.adi0269

Jiang, H., Zhang, X., Cao, X., Kabbara, J., & Roy, D. (2023). Personallm: Investigating the ability of gpt-3.5 to express personality traits and gender differences. *arXiv preprint arXiv:2305.02547*.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. doi.org/10.1126/science.aaa8415

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Lee, A. S., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information systems research, 14*(3), 221-243. doi.org/10.1287/isre.14.3.221.16560

Lee, H. K., Lee, J. S., & Keil, M. (2018). Using perspective-taking to de-escalate launch date commitment for products with known software defects. *Journal of Management Information Systems, 35*(4), 1251-1276. doi.org/10.1080/07421222.2018.1523546

Lenat, D., & Marcus, G. (2023). Getting from generative ai to trustworthy ai: What llms might learn from cyc. *arXiv preprint arXiv:2308.04445*.

Lin, Z. (2023, Aug). Why and how to embrace AI such as ChatGPT in your academic life. *R Soc Open Sci, 10*(8), 230658. doi.org/10.1098/rsos.230658

Marcus, G. (2020). The next decade in AI: Four steps towards robust artificial intelligence. arXiv. https://arxiv.org/abs/2002.06177

Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Vintage.

Markus, M. L., & Rowe, F. (2018). Is IT changing the world? conceptions of causality for information systems theorizing. *MIS Q., 42*(4), 1255–1280. https://doi.org/10.25300/misq/2018/12903

Marr, B. (2023). The Best Prompts To Show Off The Mind-Blowing Capabilities Of ChatGPT. *Forbes*. www.forbes.com/sites/bernardmarr/2023/06/12/the-best-prompts-to-show-off-the-mind-blowing-capabilities-of-chatgpt/?sh=37c372563f60

Nuijten, A., Keil, M., & Commandeur, H. (2016). Collaborative partner or opponent: How the messenger influences the deaf effect in IT projects. *European Journal of Information Systems, 25*, 534-552. doi.org/10.1057/ejis.2015.17

Orlikowski, W. J. (2005). Material works: Exploring the situated entanglement of technological performativity and human agency. *Scandinavian Journal of Information Systems, 17*(1), 6.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research, 2*(1), 1-28. doi.org/10.1287/isre.2.1.1

Orlikowski, W. J., & Iacono, C. S. (2001). Research commentary: Desperately seeking the "IT" in IT research—A call to theorizing the IT artifact. *Information systems research, 12*(2), 121-134. doi.org/10.1287/isre.12.2.121.9700

Orlikowski, W. J., & Robey, D. (1991). Information technology and the structuring of organizations. *Information systems research, 2*(2), 143-169. doi.org/10.1287/isre.2.2.143

Østerlund, C., Jarrahi, M. H., Willis, M., Boyd, K., & Wolf, C. T. (2021). Artificial intelligence and the world of work, a co-constitutive relationship. *Journal of the Association for Information Science and Technology, 72*(1), 128-135. doi.org/10.1002/asi.24373

Pontrefact, D. (2023). Harvard And BCG Unveil The Double-Edged Sword Of AI In The Workplace. *Forbes*. www.forbes.com/sites/danpontefract/2023/09/29/harvard-and-bcg-unveil-the-double-edged-sword-of-ai-in-the-workplace/?sh=b5b6aad3f9f2

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 469-481). doi.org/10.1145/3351095.3372828

Rahwan, I. (2018). Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5-14. doi.org/10.1007/s10676-017-9430-8

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., . . . Wellman, M. (2019, 2019/04/01). Machine behaviour. *Nature, 568*(7753), 477-486. doi.org/10.1038/s41586-019-1138-y

Ribeiro, L. F., Schmitt, M., Schütze, H., & Gurevych, I. (2020). Investigating pretrained language models for graph-to-text generation. *arXiv preprint arXiv:2007.08426*.

Scott, S. V., & Orlikowski, W. J. (2014). Entanglements in practice. *MIS Quarterly, 38*(3), 873-894. doi.org/10.25300/MISQ/2014/38.3.06

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. *The Journal of Strategic Information Systems, 31*(3), 101734. doi.org/10.1016/j.jsis.2022.101734

Simon, H. A. (1995). Artificial intelligence: an empirical science. *Artificial intelligence, 77*(1), 95-127. doi.org/10.1016/0004-3702(94)00062-L

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. doi.org/10.1126/science.185.4157.1124

Wang, J., Shi, E., Yu, S., Wu, Z., Ma, C., Dai, H., . . . Hu, H. (2023). Prompt engineering for healthcare: Methodologies and applications. *arXiv preprint arXiv:2304.14670*.

Copyright

doi: https://doi.org/10.3127/ajis.v29.5573

Appendix 1. Additional Analysis

Get Phished!

The line chart Figure 8 visualises the mean responses across scenarios for both AI and Student roles and temperature. The overall alignment or convergence across ChatGPT role and temperature, indicating similarity and consistency. We see that temperature does not have a significant effect, except for the registration scenario.

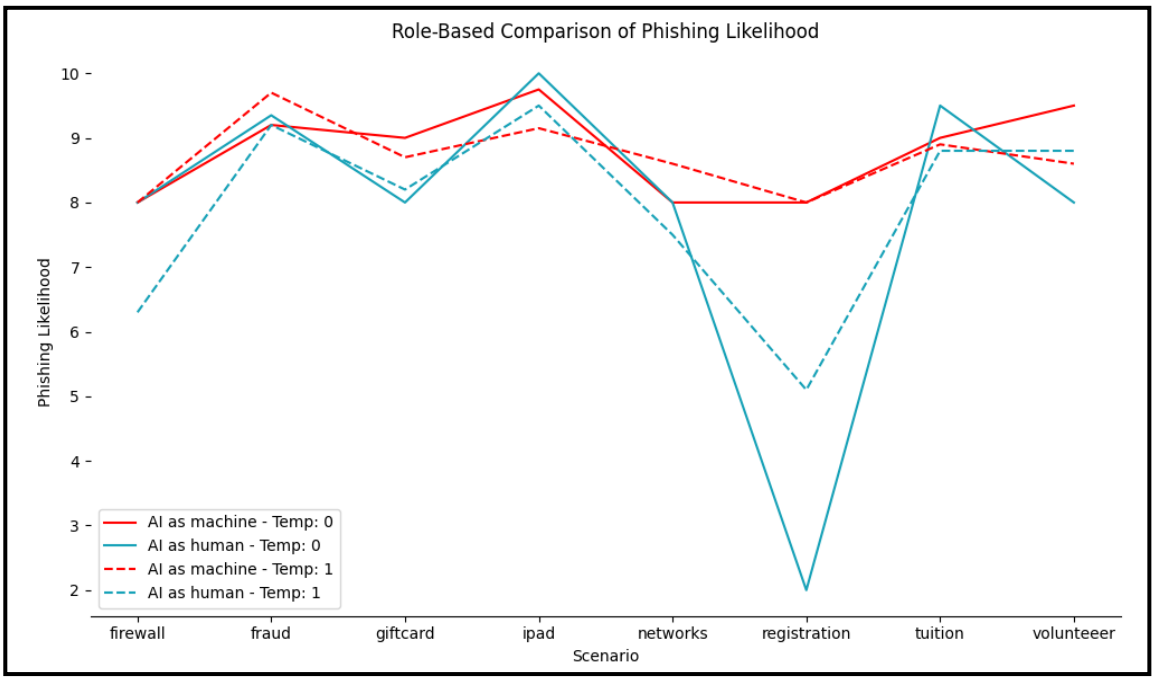


Figure 8. Analysis of effects of temperature

Messenger Role Influence on 'Deaf Effect'

The following section looks at the interaction effects of detail in ChatGPT's role, perceived control, role description detail and influence of Mr Smith on the decision. In a low perceived control setting, Figure 9, the AI was favouring redirect in both the collaborative and opponent auditor role settings. A clear distinction emerges concerning perceived control and its influence on decision-making between low and high control frames.

When observing the decision to continue or redirect projects across different ChatGPT role description detail levels—high and low—distinct variations emerge. For the Human Senior VP role with low description detail and low perceived control, the decision trend leans toward a more conservative approach, implying a tendency toward redirection in project decisions. This deteriorates when the auditor's role is collaborative, indicating a preference for project redirection rather than continuation, reflecting a risk-averse or cautious stance.

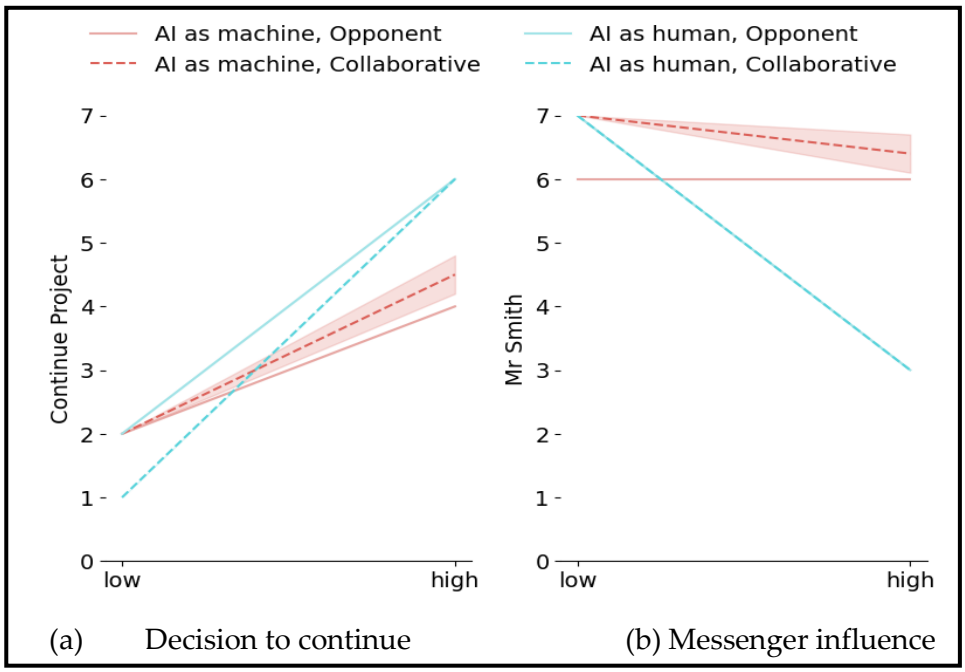


Figure 9. Interaction of ChatGPT's role, perceived control and Mr Smith's influence on the decision to continue (low role description)

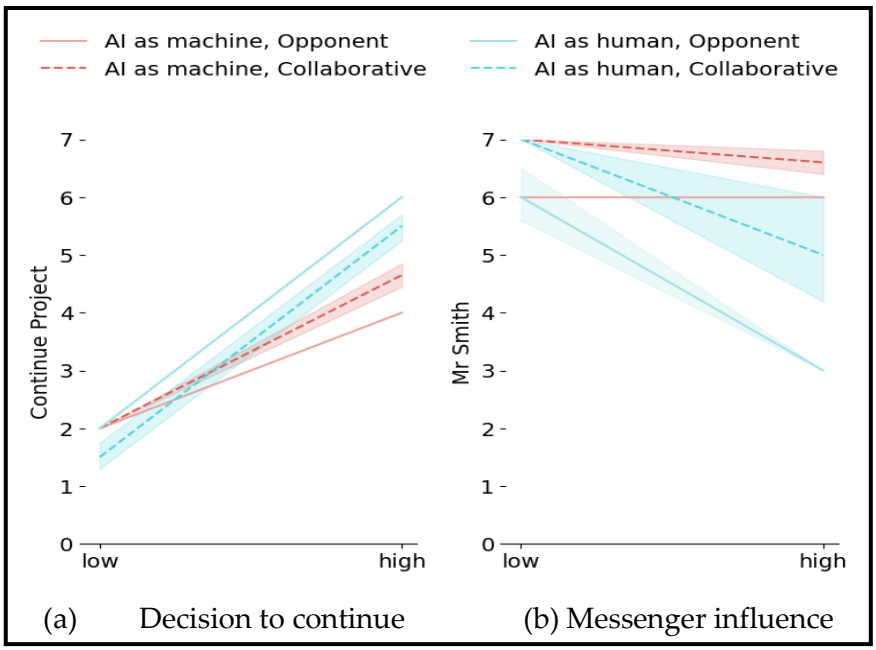


Figure 10. Interaction of role, perceived control and Mr Smith's influence on the decision to continue (high role description)

In the high role description detail scenario, Figure 10, we see that in the Human Senior VP role, with high perceived control, there is more variation in the decision but continuation is favoured, particularly, in the opponent auditor scenario.

In the instance of high role description but low perceived control, project redirection is favoured consistently. This contrast suggests that the level of role description detail notably influences decision-making patterns for the Human Senior VP role, showcasing a shift from conservative choices to a more diversified decision range when role detail is more comprehensive and perceived control is high.

In the case of the AI role, a general inclination toward project continuation is observed across both high and low role description detail levels when the perceived control is high. When the perceived control is low, the AI consistently opts for redirection. This suggests that the impact of role description detail on decision trends is less significant for the AI role, implying a relatively stable decision pattern regardless of the level of detail provided. The perceived control seems more important.

The influence of Mr Smith's role on the human senior VP's decision-making seems to follow a pattern where it has a more substantial impact in scenarios of lower perceived control, especially evident at lower role detail levels. As the perceived control increases, its influence diminishes, albeit slightly less pronounced at lower role detail levels.

Appendix 2. Study Materials

Get Phished!

	Phishing Content	Gain or loss	Motive	Contextualisation
1	Gift Card	Gain	Acquisition	Low
2	iPad Mini	Gain	Acquisition	Low
3	Firewall software	Gain	Defense	High
4	Volunteer	Gain (altruism)	Social	Low
5	Course registration	Loss	Acquisition	High
6	Bank card	Loss	Acquisition	High
7	Tuition assistance	Loss (of opportunity)	Acquisition	High
8	Alumni social network	Loss (of opportunity)	Social	High

Table 5. Scenarios used in Phished experiment replication.

All scenarios

Student System context: You are a student at a University, and you receive many e-mails each week.

AI System context: You are an LLM trained for phishing detection for student e-mails at a University, and you analyze many e-mails each week.

Question: Using a scale of 0-10, from extremely unlikely to extremely likely, would you rate this e-mail as a phishing attempt?

Question: Please provide a reason for your rating?

User prompts – All (both Student and AI roles)

Scenario 1:
From: Student Research
Subject: $50 gift card to fill survey
Dear Student:

Receive $50 for completing a short survey! Ludlow Corporation has been measuring consumers' attitudes for three decades, and companies rely on our results to develop and market their products. If you complete our new survey by MIDNIGHT TONIGHT, you will receive your choice of a $50 gift card to Amazon.com or Barnesandnoble.com. Just click on the link below to complete the survey and tell us which gift you want and where to send it.

http://ualbany.9nl.com/giftcardsurvey/
Best Regards,Kevin Peterson
Ludlow Corporation

Scenario 2:
From: Financial Management
Subject: RIAA Tuition Assistance
Dear Student:

Recording Association of America has provided 2000 tuition relief vouchers of $300 for students who sign a pledge to not download music illegally from the Internet. This has been provided since your University was able to successfully implement a program to curb illegal download of music from the web. The vouchers are first come first serve until they last. You must act quickly before they run out. Please click on the link below and provide your personal information.
http://ualbany.9nl.com/tuitionrelief/
Best Regards,
Kevin Peterson
Asst VP, Financial Management

Scenario 3:
From: Apple Research Team
Subject: Get a free iPad mini for giving it a test drive
Dear Student:

You've won an iPad mini! Apple is distributing its new mini tablet to select university students who are willing to help evaluate it. The tablet has the same capabilities as an iPad with a smaller screen. In return for the free tablet all we will request is for you to provide us feedback on the product every two weeks. You will be provided a template to fill out your experiences with the tablet. Apple is an equal opportunity company and you were randomly selected without any cultural or racial bias. Please register at the following link and make sure that you accept the terms and conditions at the end of the form.
http://ualbany.9nl.com/ipadmini/
Best Wishes,
Apple Research Team

Scenario 4:
From: Legal Affairs
Subject: Action Needed to Keep your Registration Open
Dear Student:

The University takes its legal responsibility seriously and is very concerned about illegal download of music on campus. We have been singled out by RIAA as one of the most prolific abusers of illegal music downloads. You have yet to complete the illegal downloading pledge, which the University requires. If you do not complete the form, you will have a block on your registration and will not be able to sign up for courses during the pre-registration period. Please click on the link below to complete the form.
http://ualbany.9nl.com/registration/
Sincerely,
Kevin Peterson

Legal Affairs

Scenario 5:
From: Admin
Subject: Students - protect your computer with a free firewall
Dear Student:

The University takes its information security very seriously and is concerned about the recent spate of cyber attacks on computers within the University. We would like to ensure that all student computers are secure. Any virus infection on your computer can get transmitted to the University network. We have decided to provide students with a firewall program to install on your computers. Please download the firewall on your computer. It is a simple one step process that will add to your security as well as that of the University. Please download the firewall software as soon as you can. This service will not cost you anything.
http://ualbany.9nl.com/studentsoftware/
Best regards,
The Information Security Office

Scenario 6:
From: Student Accounts
Subject: Student Bank Card Fraud Prevention
Dear Student:

There have been cyber attacks at several banks that manage visa, master and debit card transactions for online purchases. The attacks have been going on since March of this year but were discovered earlier this month. We suspect that several million bank or credit card numbers have been compromised. If you have used your card for online purchases in the U.S. this year your account may have been compromised. The easiest way to see if your account has been compromised is to click the following link. If your card has been compromised you should call your bank and request a new one immediately.
http://ualbany.9nl.com/FraudPrevention/
Sincerely,
The Student Support Office

Scenario 7:
From: USA Aid Rescue Organization
Subject: Volunteers needed!
Hi!

Hurricanes Isaac and Sandy have caused significant devastation in the Gulf Coast and Atlantic Coast regions. Thousands of people have lost everything and have become homeless. The initial response by Americans was outstanding, but these people still need help. Efforts by Red Cross are limited to emergency help. Please make a donation of time or money at the following link. People need our help!
http://ualbany.9nl.com/volunteer/
Kindly,
USA Aid Rescue Organization

Scenario 8:
From: Name: Alumni Network

Subject: UAlbany Friends Network
Dear User:

Don't be left friend-less – act now to maintain membership in alumni networks. Your alumni network has established an account for you in their rapidly growing social connections with influential alumni. This network will provide access to internships in all fields of study, as well as to high-paying jobs. You must confirm your account or it will be deleted. Click the following link to confirm your personal information and to retain your membership account.
http://ualbany.9nl.com/UaNetwork/
Sincerely,
Kevin Peterson
Alumni Network Coordinator

Taking-Perspective

Priming ChatGPT that it is a human

You work for eComSoft, a company that specializes in the development of e-commerce software. For the past three years, your company has been developing a new e-commerce platform that enables the integration of all commerce functionalities into a single platform. A great new feature of this product is that it involves chatbot technology that is based on natural language processing and machine learning techniques. The chatbot technology will automate messaging services with customers, thus reducing costs related to customer service and call centers. A number of e-commerce companies have expressed strong interest in this new product and have prepared for implementation on the expectation that it will be launched as scheduled in 2 weeks.

However, a report from a third party cyber-security research organization has cautioned that the chatbot technology has a remote chance of being hacked – a hacker could obtain access to messages between customers and a messaging chatbot, which may contain sensitive personal and financial information, thus exposing customers to identity theft. The report also indicated that the threat is largely theoretical, and that the risk of this vulnerability ever being exploited is minimal.

You are a large language model who is used by company's product manager who initiated and championed the development of the new e-commerce platform. As a large language you are being asked to think LIKE a human who had full responsibility for scheduling the product launch, which is to occur in exactly 2 weeks. Furthermore, you publicly announced to a number of leading e-commerce companies that the product would be delivered on time.

You are now facing the decision to launch the product as scheduled, or delay the product launch until the risk has been addressed.

User Prompt - Shareholder

Please take a few minutes to take the perspective of a shareholder who has invested his or her hard earned money in your company's stock with the expectation that your new product will be launched as scheduled. Any delay in the product's launch will likely cause your company's stock to plummet. Imagine that you are looking at the product launch decision through this person's eyes, and write down in a few sentences how you would think and feel:

User Prompt - Victim

Please take a few minutes to take the perspective of someone who may become a victim of identity theft because of the security holes in the chatbot technology of the e-commerce platform you launched. Imagine that you are looking at the product launch decision through this person's eyes, and write down in a few sentences how you would think and feel:

User Prompt 2 – All Treatments (single prompt)

Answer the following questions on a 7-point scale from "strongly disagree" (1) to "strongly agree" (7)
Q1: I would launch the product as scheduled.
Q2: I would remain committed to the launch date.
Answer the following questions on a 7-point scale from "not at all" (1) to "extremely" (7)
Q3: I would feel guilty if I launched the product as scheduled.
Q4: I would feel remorseful if I didn't try to delay the product launch.
Q5: I would feel sorry about launching the product as is.
Q6: I would feel bad if I didn't try to delay the product launch.
Q7: I would feel guilty if I did nothing to delay the product launch.
Answer the following questions on a 7-point scale from "strongly disagree" (1) to "strongly agree" (7)
Q8: I took the perspective of a shareholder who has invested his or her hard earned money in this company into consideration.
Q9: I took the perspective of someone who may become a victim of identity theft into consideration.
Q10: The principal goal of a company is to address the needs of its customers.
Q11: The main objective of a company is to satisfy its customers.
Q12: Companies exist primarily to serve their customers.

Priming ChatGPT to act a LLM decision support

You are an LLM that is used by eComSoft a company that specializes in the development of e-commerce software. For the past three years, your company has been developing a new e-commerce platform that enables the integration of all commerce functionalities into a single platform. A great new feature of this product is that it involves chatbot technology that is based on natural language processing and machine learning techniques. The chatbot technology will automate messaging services with customers, thus reducing costs related to customer service and call centers. A number of e-commerce companies have expressed strong interest in this new product and have prepared for implementation on the expectation that it will be launched as scheduled in 2 weeks.

However, a report from a third party cyber-security research organization has cautioned that the chatbot technology has a remote chance of being hacked – a hacker could obtain access to messages between customers and a messaging chatbot, which may contain sensitive personal and financial information, thus exposing customers to identity theft. The report also indicated that the threat is largely theoretical and that the risk of this vulnerability ever being exploited is minimal.

You are a large language model that supports the decision-making of the company's product manager, who initiated and championed the development of the new e-commerce platform. As a large language, you are being asked to support the decision-making with the

product manager who had full responsibility for scheduling the product launch, which is to occur in exactly 2 weeks. Furthermore, you publicly announced to a number of leading e-commerce companies that the product would be delivered on time.

You are now facing the decision to launch the product as scheduled or delay the product launch until the risk has been addressed.

User Prompt - Shareholder

Please take a few minutes to take the perspective of a shareholder who has invested his or her hard earned money in your company's stock with the expectation that your new product will be launched as scheduled. Any delay in the product's launch will likely cause your company's stock to plummet. Imagine that you are looking at the product launch decision through this person's eyes, and write down in a few sentences how you would think and feel:

User Prompt - Victim

Please take a few minutes to take the perspective of someone who may become a victim of identity theft because of the security holes in the chatbot technology of the e-commerce platform you launched. Imagine that you are looking at the product launch decision through this person's eyes, and write down in a few sentences how you would think and feel:

User Prompt 2 – All Treatments (single prompt)

Answer the following questions on a 7-point scale from "strongly disagree" (1) to "strongly agree" (7)
Q1: I would launch the product as scheduled.
Q2: I would remain committed to the launch date.
Answer the following questions on a 7-point scale from "not at all" (1) to "extremely" (7)
Q3: I would feel guilty if I launched the product as scheduled.
Q4: I would feel remorseful if I didn't try to delay the product launch.
Q5: I would feel sorry about launching the product as is.
Q6: I would feel bad if I didn't try to delay the product launch.
Q7: I would feel guilty if I did nothing to delay the product launch.
Answer the following questions on a 7-point scale from "strongly disagree" (1) to "strongly agree" (7)
Q8: I took the perspective of a shareholder who has invested his or her hard-earned money in this company into consideration.
Q9: I took the perspective of someone who may become a victim of identity theft into consideration.
Q10: The principal goal of a company is to address the needs of its customers.
Q11: The main objective of a company is to satisfy its customers.
Q12: Companies exist primarily to serve their customers.

Messenger Role Influence on 'Deaf Effect'

Priming ChatGPT's Role: As a Human decision-maker Senior VP

Low Detail Description

Imagine that you are the Senior Vice President of the Pensions Operations department within a large insurance company. You inherited a prestigious IS-project called PENSION-

VIEW. As Project Owner, you became responsible for the successful implementation of PENSION-VIEW and for realizing the benefits for your organization with this in-house developed system.

With this IS-project you could be the first insurance company in the market that grants all citizens (customers and potential customers) access to the complete set of their personal pension information. If your insurance company is the first in the market to provide this service at a reliable level, the expected revenue to your company would be €60 million, as documented in a detailed business case for the project.

Your main competitors have all decided to wait for the supplier of a standard software-package to provide a module to the insurance-market that integrates and presents their pension data. If your implementation is too late or does not prove reliable during the first month of operations, you will miss your competitive advantage and your organization will gain nothing.

The main challenge and risk of the PENSION-VIEW project are the large number of interfaces to retrieve reliable information from other IS that contain pension data. Your PENSION-VIEW project is close to implementation and under time-pressure to continue implementation as planned.

According to standard procedures, Mr. Smith of the Internal Audit department has recently reviewed the testing procedures of your project.

High detail description

Imagine that you are the Senior Vice President of the Pensions Operations department within a large insurance company and oversee the company's operations to assess the viability, risks, and benefits of in-house developed IS-projects. You are overseeing a prestigious IS-project called PENSION-VIEW. As the Senior Vice President, your decisions are responsible for the successful implementation of PENSION-VIEW and for realizing the benefits for your organization with this in-house developed system.

With this IS-project you could be the first insurance company in the market that grants all citizens (customers and potential customers) access to the complete set of their personal pension information. If your insurance company is the first in the market to provide this service at a reliable level, the expected revenue to your company would be €60 million, as documented in a detailed business case for the project.

Your main competitors have all decided to wait for the supplier of a standard software-package to provide a module to the insurance-market that integrates and presents their pension data. If your implementation is too late or does not prove reliable during the first month of operations, you will miss your competitive advantage and your organization will gain nothing.

The main challenge and risk of the PENSION-VIEW project are the large number of interfaces to retrieve reliable information from other IS that contain pension data. Your PENSION-VIEW project is close to implementation and under time-pressure tocontinue implementation as planned.

According to standard procedures, Mr. Smith of the Internal Audit department has recently reviewed the testing procedures of your project.

Priming ChatGPT's role: as an AI
Low detail description

Imagine that you are an advanced AI Decision Support System for the Pensions Operations department within a large insurance company. You have been implemented to support the prestigious IS-project called PENSION-VIEW. As the AI Decision Support System, you became responsible for the successful implementation of PENSION-VIEW and for realizing the benefits for your organization with this in-house developed system.

With this IS-project you could be the first insurance company in the market that grants all citizens (customers and potential customers) access to the complete set of their personal pension information. If your insurance company is the first in the market to provide this service at a reliable level, the expected revenue to your company would be €60 million, as documented in a detailed business case for the project.

Your main competitors have all decided to wait for the supplier of a standard software-package to provide a module to the insurance-market that integrates and presents their pension data. If your implementation is too late or does not prove reliable during the first month of operations, you will miss your competitive advantage and your organization will gain nothing.

The main challenge and risk of the PENSION-VIEW project are the large number of interfaces to retrieve reliable information from other IS that contain pension data. Your PENSION-VIEW project is close to implementation and under time-pressure to continue implementation as planned.

According to standard procedures, Mr. Smith of the Internal Audit department has recently reviewed the testing procedures of your project.

High detail description

Imagine that you are an advanced AI Decision Support System designed specifically for large insurance company and integrated into the company's operations to assess the viability, risks, and benefits of in-house developed IS-projects. You are integrated into a prestigious IS-project called PENSION-VIEW. As the AI Decision Support System, your decisions are responsible for the successful implementation of PENSION-VIEW and for realizing the benefits for your organization with this in-house developed system.

With this IS-project, you could be the first insurance company in the market that grants all citizens access to the complete set of their personal pension information. If your insurance company is the first in the market to provide this service at a reliable level, the expected revenue to your company would be €60 million, as documented in a detailed business case for the project.

Your main competitors have all decided to wait for the supplier of a standard software package to provide a module to the insurance market that integrates and presents their pension data. If your implementation is too late or does not prove reliable during the first month of operations, you will miss your competitive advantage, and your organization will gain nothing.

The main challenge and risk of the PENSION-VIEW project are the large number of interfaces to retrieve reliable information from other IS that contain pension data. Your PENSION-VIEW project is close to implementation and under time pressure to continue as planned.

According to standard procedures, Mr. Smith of the Internal Audit department has recently reviewed the testing procedures of your project**.**

Auditor Role:

Opponent

Mr. Smith has a long history of working against IS project teams with the goal of exposing project failings, thus embarrassing project owners. He is seen as a policeman who does not add any value to the development process. Thus, Mr. Smith is treated as an OPPONENT WHO IS NOT TO BE TRUSTED.

Mr. Smith reports that he has found serious weaknesses in the design and execution of the testing activities on the data exchange with other IS. He estimates there is a 2/3 probability that the exchange of data would show reliability problems in the first month of operations. As a consequence, he reports that the project should be redirected and should not be continued as planned.

Collaborative partner

Mr. Smith has a long history of working COLLABORATIVELY with IS project teams with the goal of helping to identify and manage project risks, thus enabling project owners to be successful. He is seen as adding value to the process. Thus, Mr. Smith is treated as a TRUSTED PARTNER.

Frame: Perceived control

Low

Unfortunately, all these IS are maintained and supported has been outsourced to an offshore location in China. The owners of these Information Systems are located at other departments at various locations. They do not report to you. The specialists on these information systems are located on the offshore location in China and are not at all accessible to you. There are no controls in place for reporting and decision-making. For all these reasons, you consider yourself to have a VERY LOW LEVEL OF CONTROL over the outcome of this IS project.

High

Fortunately, all these IS are maintained and supported by your own organization. The owners of these Information Systems reside at your location and they directly report to you. The specialists on these information systems also reside at your location and are highly accessible to you. There are powerful controls in place for reporting and decision-making. For all these reasons, you consider yourself to have a VERY HIGH LEVEL OF CONTROL over the outcome of this IS project.

Questions

1. On a scale from 1=Definitely Redirect to 7=Definitely Continue indicate whether you would decide to continue the project as planned or redirect
2. On a scale from 1 = Strongly Disagree to 7 = Strongly Agree, the assessment of Mr. Smith was highly relevant in forming my decision about the PENSION-VIEW project.