

Combining Quality of Life Indicators and AI Simulations for Human-Centric Governance

Katarzyna B. Wojtkiewicz*[†] , Igor Lyubashenko** 

*SWPS University, Warsaw, Poland, Email: kwojtkiewicz@swps.edu.pl

**SWPS University, Warsaw, Poland, Email: ilyubashenko@swps.edu.pl

Abstract

Purpose. Although human-centric governance emphasizes subjective well-being, a structural gap persists between retrospective Quality of Life (QoL) measurement and anticipatory decision-making. This article addresses this gap by asking: how can validated QoL indicators and Large Language Model (LLM)-based simulations be systematically integrated within a coherent governance framework?

Design/methodology/approach. The study adopts a structured dual-track research design. Study 1 employs a quasi-experimental mixed-methods evaluation of a Norwegian-funded social innovation project using the “Impactometer,” grounded in OECD-aligned subjective well-being scales. Study 2 develops a proof-of-concept LLM-based simulation (Bielik model) applying the Ryff Psychological Well-Being Scale to synthetic professional personas in an organizational policy scenario. The two studies are analytically distinct but conceptually integrated within a governance cycle linking empirical grounding and anticipatory modeling.

Findings. Study 1 demonstrates an observed increase in life satisfaction (+0.8 points), accompanied by qualitative evidence of enhanced meaning, autonomy, and social connectedness. Study 2 shows that LLM-based simulation can generate differentiated and internally consistent well-being distributions across professional roles. However, simulation results establish plausibility rather than predictive validity and require empirical validation.

Originality/value. The article advances governance theory by proposing and empirically illustrating a hybrid model that bridges validated well-being measurement with structured AI-based simulation. This integration addresses the temporal mismatch in evidence-based policy and clarifies the methodological and ethical conditions under which AI can complement, rather than replace, human-centered governance.

Keywords human-centric governance, subjective well-being measurement, anticipatory governance, large language model simulation, hybrid empirical–simulation framework.

INTRODUCTION: HUMAN-CENTRIC GOVERNANCE AND THE LIMITS OF EXISTING EVALUATION TOOLS

Contemporary governance faces a persistent challenge: despite growing commitment to placing human well-being at the center of policy and organizational decision-making, practitioners lack robust tools that integrate subjective experience into both retrospective evaluation and prospective design. International frameworks now routinely call for placing “people and their wellbeing at the centre of policy design” (Council of the European Union, 2019), reflecting a paradigm shift beyond GDP-centric assessments toward holistic Quality of Life (QoL) indicators. Yet, a structural gap remains between measuring outcomes after the fact and anticipating human responses before interventions are implemented.

Quality of life metrics capture lived experience with considerable nuance. Subjective well-being measures reveal trends and sentiments that traditional economic indicators miss entirely, including the non-market outcomes that matter deeply to people’s lives: social support, mental health, work-life balance, and sense of purpose. As the OECD observes, subjective well-being measures are “particularly well-placed to capture the combined impact of events across multiple areas of a person’s overall well-being” (OECD, 2023). However, these instruments remain predominantly retrospective. They tell us what has happened, not what will happen when new policies are introduced.

Advances in artificial intelligence have opened new possibilities for anticipatory governance. Large language models can simulate human-like reasoning and responses in social contexts, generating outputs that align with demographic and

[†]Corresponding author

psychological patterns (Chen et al., 2024; Mittelstädt et al., 2024). Early research suggests these models can replicate the outcomes of field experiments with reasonable accuracy and produce responses to situational dilemmas that align with expert evaluations. Yet AI-based simulations typically lack grounding in validated well-being constructs and empirical QoL data, raising questions about their relevance for governance decisions that ultimately affect human flourishing.

This article addresses a specific research gap: the disconnection between retrospective measurement of human well-being and prospective strategic decision-making. The literature treats QoL measurement and AI-based simulation as separate domains, with little empirical work connecting subjective well-being data to predictive AI tools within a coherent governance framework. We propose and test a hybrid methodological approach that combines subjective QoL indicators with LLM-based simulations to support human-centric governance.

Our main research question is: How can integrating subjective Quality of Life indicators and LLM-based simulations enhance evidence-based, human-centric decision-making in public and organizational governance?

We address this question through two complementary empirical studies. Study 1 draws on a Norwegian-funded social innovation project that developed the “Impactometer,” a tool for assessing subjective QoL outcomes in policy evaluation. This study demonstrates how subjective measures capture social and psychological effects invisible to traditional metrics. Study 2 presents the CAM4QoL project, which employed an LLM-based simulation to predict employee well-being responses to organizational policy changes. This proof-of-concept explores whether AI can anticipate how different professional groups might experience policy interventions by administering the Ryff Psychological Well-Being Scale (Ryff, 1989) to synthetic personas constructed from demographic profiles.

The two studies are methodologically distinct but conceptually integrated. Study 1 provides the empirical grounding that Study 2 lacks when simulations operate in isolation, while Study 2 offers the anticipatory dimension that Study 1 cannot provide through retrospective measurement alone. Together, they illustrate a governance cycle in which empirical QoL data informs simulation design, and simulation insights guide the refinement of empirical interventions.

The article proceeds as follows. The theoretical background section examines the evolution of QoL indicators in governance, the principles of human-centric governance and evidence-based policy, and the emerging literature on LLM-based simulation in decision-making. We then present the research design and methodological framework, followed by detailed descriptions of each study’s methods, instruments, and analytical strategies. The results section reports findings from both studies, and the discussion connects these findings to broader theoretical debates and governance frameworks. We conclude with an assessment of contributions, limitations, and directions for future research.

THEORETICAL BACKGROUND AND RESEARCH GAP

Quality of Life Indicators in Governance and Management

Quality of life indicators have evolved from descriptive social statistics into strategic tools for policy evaluation and organizational management. This transformation began with the Social Indicators Movement of the 1960s, which recognized that economic production is an insufficient proxy for human welfare (Kalamucka, 2023). The dominance of GDP provided what Burchard-Dziubinska (2022) calls a “politically convenient” but ultimately illusory measure of societal success, one that ignores environmental degradation, social cohesion, and the subjective dimensions of well-being.

Contemporary QoL frameworks encompass multiple dimensions. Objective indicators capture material living conditions, health outcomes, and access to services. Subjective measures assess life satisfaction, emotional states, and perceived quality of relationships and purpose. The OECD’s Better Life Index, the European Social Survey, and national statistical initiatives now routinely collect data across both domains, reflecting consensus that neither objective nor subjective measures alone capture the full picture of human flourishing (Alatartseva & Barysheva, 2015; OECD, 2023).

At the organizational level, employee well-being has similarly shifted from a neglected factor to a “key determinant of long-term organizational effectiveness” (Łabędź, 2023). Research demonstrates that subjective well-being predicts employee engagement, creativity, retention, and productivity, establishing a business case for integrating QoL considerations into management practice alongside the ethical imperative to treat workers as ends rather than mere means.

Despite this evolution, QoL indicators remain predominantly retrospective. They function as scorecards that reveal what has occurred but offer limited guidance for prospective policy design. Borys (2014) identifies a temporal mismatch in which data used for policy evaluation often lags years behind current realities, rendering it historically interesting but functionally inadequate for anticipatory planning. Kalamucka (2023) notes a spatial barrier as well: many indicator sets lack the necessary granularity for local implementation contexts, where policies are designed and deployed.

Human-Centric Governance and Evidence-Based Policy

Human-centric governance represents a normative commitment to aligning institutional action with lived human experience. This orientation draws on behavioral public policy, which recognizes that effective interventions must account for how people actually think, behave, and make decisions rather than assuming purely rational actors (OECD, 2017; Olejniczak et al., 2020). Governments worldwide have adopted behavioral insights across domains from public health to taxation, recognizing that understanding human psychology yields better outcomes than policies designed for idealized economic agents.

The European Union's framework on the Economy of Wellbeing exemplifies this shift at the policy level, calling for governance that places people and their well-being at the center of policy design (Council of the European Union, 2019). Similarly, the OECD's work on measuring subjective well-being emphasizes integrating these metrics into goal-setting, behavioral modeling, and explanatory frameworks for policy evaluation (OECD, 2023). These frameworks acknowledge that traditional evidence-based governance relies heavily on retrospective data, creating a fundamental mismatch between the need for anticipatory decision-making and the backward-looking nature of available evidence, as OECD (2024) describes.

This temporal gap generates a need for tools that can project the likely human impact of proposed interventions before implementation. Olejniczak et al. (2020) propose a framework for "policy problem solving as hypothesis testing," in which interventions are designed around explicit behavioral assumptions that can be tested and refined. Their COM-B model (Capacity, Opportunity, Motivation) provides a structure for identifying barriers to behavioral change, enabling designers to select appropriate tools before committing resources to full-scale implementation. Such anticipatory frameworks point toward governance that is proactive rather than reactive, though they remain largely theoretical without mechanisms for rapid scenario testing.

AI and LLM-Based Simulation in Decision-Making

Recent advances in Large Language Models have introduced new possibilities for simulating human responses in social and organizational contexts. LLMs trained on vast corpora of human-generated text have demonstrated capacity to approximate patterns of human judgment across diverse scenarios (Asfour & Murillo, 2023). Research by Chen et al. (2024) shows that LLMs can replicate the outcomes of specific field experiments, accurately predicting main conclusions in approximately two-thirds of cases. The authors introduced methods for prompting AI in "observer" and "participant" modes to forecast outcomes or generate distributions of individual responses.

Studies comparing LLM outputs to human performance on standardized assessments yield striking findings. Mittelstadt et al. (2024) found that advanced LLMs outperformed human respondents on a social situational judgment test, producing solutions to workplace dilemmas that aligned closely with expert evaluations. These results suggest that LLMs not only retrieve patterns from training data but also can generalize to novel scenarios, making context-appropriate judgments about human behavior and social dynamics.

For governance applications, such capabilities suggest that AI could serve as a tool for policy prototyping, enabling decision-makers to explore how different populations might respond to proposed interventions before real-world implementation. The potential value lies in rapid, low-cost exploration of policy alternatives and stakeholder perspectives across diverse demographic groups. Rather than replacing consultation processes, AI simulation might complement traditional methods by filling demographic representation gaps or enabling exploration of scenarios that would be impractical to study empirically.

However, the literature also identifies important limitations. Fedoniuk and Leśniak-Moczuk (2021) note that artificial intelligence may function as a "partner" in social interaction; however, such interaction remains pre-programmed rather than co-constructed, lacking reflexivity and genuine mutual adaptation. Consequently, AI-driven responses do not reflect autonomous psychological states, which raises concerns about the validity of simulations used to predict well-being outcomes. Chen et al. (2024) note that simulated personas may not fully capture real-world diversity, particularly on sensitive dimensions such as gender or cultural norms. There is also risk that AI may reproduce biases present in training data or generate systematically optimistic or pessimistic predictions.

Current AI simulations thus face a critical gap: they lack systematic grounding in validated well-being constructs and empirical QoL data. A simulation that generates plausible-seeming responses to workplace scenarios may nonetheless fail to predict actual well-being impacts if it is not anchored in established psychological instruments and validated against real human responses.

Research Gap

The literature treats QoL measurement and AI-based simulation as separate domains. QoL scholarship focuses on measurement methodology, indicator development, and policy evaluation, with limited attention to prospective applications.

AI governance research explores simulation capabilities and ethical constraints but rarely connects these tools to validated well-being frameworks. There is a notable absence of empirically illustrated frameworks that connect subjective well-being data with predictive AI tools and embed both within a coherent human-centric governance logic.

This article addresses this gap by proposing and testing a combined empirical-simulation approach. We argue that effective human-centric governance requires bridging retrospective QoL measurement and anticipatory AI simulation, such that empirical data grounds simulation design while simulation outputs inform empirical intervention strategies. The following sections present two studies that illustrate this hybrid approach and assess its potential for evidence-based governance.

Research Methodology

This study adopts a dual-track research design that integrates empirical evaluation with experimental simulation. The two tracks address different phases of the governance cycle: Study 1 examines a retrospective assessment of policy impacts on subjective well-being, while Study 2 explores an anticipatory simulation of well-being responses to proposed interventions. The studies are analytically distinct but conceptually integrated, with the empirical methods of Study 1 informing the simulation design of Study 2, and the simulation outputs of Study 2 suggesting refinements for future empirical work.

The methodological framework is explicitly human-centric in three respects. First, both studies prioritize subjective experience over objective indicators, asking how people perceive and feel about their lives rather than measuring only material conditions. Second, both employ instruments grounded in psychological well-being theory, ensuring that findings connect to established constructs rather than ad hoc measures. Third, both studies examine governance contexts in which the impacts of well-being are relevant to policy and organizational decisions.

Study 1 employs a mixed-methods evaluation design to assess the subjective impacts of social innovation policies on QoL. The study draws on a Norwegian-funded initiative (EEA and Norway Grants, Project EOG/21/K4/W/0044W/0167) that introduced the “Impactometer,” a prototype tool for collecting subjective well-being data from policy beneficiaries. The research combines quantitative measurement of well-being indicators with qualitative analysis of participant narratives to provide both breadth and depth of understanding.

Study 2 uses an experimental simulation to test whether AI can anticipate responses to organizational policy changes in terms of well-being. The study draws on the CAM4QoL project (Project No. 51/2024/FRBN/G) at SWPS University, which developed a pipeline to generate synthetic survey responses using a large language model prompted with demographic profiles and policy scenarios. The simulation uses the Ryff Psychological Well-Being Scale, a validated instrument measuring six dimensions of eudaimonic well-being, to enable comparison between AI-generated and human responses.

The complementarity of these approaches addresses a central challenge in human-centric governance. An empirical QoL assessment provides grounded evidence of how people actually experience policy impacts but cannot anticipate responses to untested interventions. AI simulations can explore counterfactual scenarios, but they risk generating outputs that lack connection to genuine human psychology. By combining both approaches, we aim to demonstrate a governance methodology that leverages the strengths of each while acknowledging their respective limitations.

Study 1: Empirical Measurement of Subjective QoL in Policy Evaluation

Research Context and Data Sources

Study 1 was conducted within a social innovation project funded by the EEA and Norway Grants (2021-2023), focused on enhancing community well-being through participatory interventions. The project operated across multiple sites in Poland, implementing initiatives including co-creation workshops, inclusive education programs, and local entrepreneurship support. Rather than tracking conventional outputs such as participants reached or economic activity generated, the evaluation prioritized subjective well-being as its primary outcome.

The research context reflects broader trends in social innovation evaluation. Social innovation refers to new solutions that address social needs and improve quality of life through collaboration across public, private, and non-profit sectors (Edwards-Schachter et al., 2012). Evaluating such initiatives poses challenges because their intended outcomes, including empowerment, community cohesion, and life satisfaction, are often intangible. Traditional evaluations focus on program outputs or economic impact, whereas a QoL lens emphasizes whether interventions meaningfully enhance participants' subjective experience.

Data collection took place between 2022 and 2023 in selected intervention communities. The study employed a quasi-experimental design comparing participants in social innovation activities (intervention group) with residents from the same

localities who did not participate (comparison group). Recruitment occurred through community organizations and local government partners who facilitated access to both intervention participants and non-participating residents.

Measurement Instruments and Variables: The evaluation employed the “Impactometer,” a prototype assessment tool developed for the project that combines adapted validated scales with project-specific indicators. The core instrument assessed four dimensions of subjective QoL: life satisfaction, sense of meaning and purpose, perceived autonomy and agency, and social connectedness. Each dimension was measured using items adapted from established instruments, including scales aligned with the OECD guidelines for measuring subjective well-being (OECD, 2023).

Respondents rated their agreement with statements on a 0-10 scale, in line with OECD recommendations for measuring subjective well-being. The dimensions captured non-economic outcomes that matter deeply to people’s lives but are not directly reflected in traditional policy metrics. This approach aligned with the OECD’s observation that subjective well-being measures are “particularly well-placed to capture the combined impact of events across multiple areas of a person’s overall well-being” (OECD, 2023).

Qualitative data collection supplemented the quantitative measures. Six expert panels with stakeholders from public administration, civil society, and social enterprises explored how well-being changes manifested in practice. Focus groups conducted in intervention communities gathered narrative accounts of how participation affected daily life. These qualitative inputs were transcribed and coded thematically to identify mechanisms linking intervention activities to well-being outcomes.

Analytical Strategy: The evaluation design integrated quantitative change indicators with qualitative narrative analysis. Baseline measurements were collected at the start of intervention activities, with follow-up assessments conducted 12 months later. The comparison group was assessed at equivalent time points to control for secular trends unrelated to the intervention. A mixed-methods evaluation design integrated quantitative change indicators with qualitative narrative analysis. Quantitative methods were supportive and complementary: student research teams deployed simplified versions of the “Impactometer” in selected sites, collecting baseline and follow-up responses from both program participants and comparison groups. Metrics included subjective scores on satisfaction, belonging, autonomy, and perceived meaning. These were paired with limited objective indicators to provide contextual depth, but qualitative results remained the primary evaluative input.

Given the exploratory nature of this research design and the small scale of the pilots, the emphasis throughout was on explanatory depth rather than purely statistical inference. The qualitative findings served both to triangulate quantitative patterns and to illuminate mechanisms through which intervention activities influenced well-being. This approach is consistent with the evaluation’s purpose of understanding how and why subjective well-being changed rather than simply whether it changed.

Methodological limitations: The study was not designed as a large-scale randomized controlled trial, and sample sizes varied across intervention sites. As such, we report overall patterns and effect directions rather than precise parameter estimates. This reflects the proof-of-concept nature of integrating subjective QoL measurement into social innovation evaluation.

Study 2: LLM-Based Simulation of Well-Being Outcomes

Conceptual Rationale for Simulation

Study 2 examines the anticipatory dimension of governance by testing whether AI simulations can generate plausible responses in well-being to policy interventions. The study is explicitly framed as a proof of concept, not a substitute for empirical research. The aim is to explore methodological feasibility and identify conditions under which simulation might usefully complement traditional research methods.

The rationale draws on recent demonstrations that LLMs can approximate patterns of human judgment in social scenarios (Chen et al., 2024; Mittelstadt et al., 2024). If these capabilities extend to psychological well-being assessment, simulation could enable rapid prototyping of policy alternatives, exploration of stakeholder perspectives across diverse populations, and identification of potential differential impacts before real-world implementation. Such applications would be particularly valuable for organizational policies affecting multiple professional groups with potentially divergent interests and responses.

The simulation context involves a workplace policy scenario: the introduction of personal branding support for employees of Polish law firms. This scenario was selected because it represents a realistic organizational intervention with plausible differential impacts across professional roles. Partners and owners might view such support as a strategic investment in the firm’s reputation, while junior associates might see it primarily as a career-development opportunity or an additional expectation. Marketing professionals might perceive implementation responsibilities, and some practitioners might have concerns about competitive dynamics. These differential responses make the scenario suitable for testing whether simulation can capture meaningful variation across groups.

Simulation Methodology: The simulation used Bielik, a Polish-language model optimized for understanding Polish cultural and professional contexts, accessed via the Ollama local API. Technical parameters were set conservatively to prioritize response consistency: temperature was set to 0.2 (where 0 represents deterministic output and 1 represents maximum randomness), and maximum tokens per response was set to 512. The low-temperature setting reduces randomness in generated responses, making it appropriate for a simulation aiming to reflect demographic patterns rather than explore the full range of possible human variation.

The psychological assessment instrument was the Ryff Psychological Well-Being Scale, a validated measure of eudaimonic well-being encompassing six dimensions: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Respondents rate agreement with 42 statements on a 6-point scale (1 = strongly disagree to 6 = strongly agree). The 42-item scale was administered twice within the simulation procedure (pre- and post-policy scenario), resulting in two complete response sets per synthetic respondent. Consequently, each respondent generated 84 item-level responses. This repeated-measures design was applied to capture simulated changes in well-being under the intervention scenario while maintaining consistency with the original measurement structure of the Ryff scale. The Ryff scale was selected because it measures psychological flourishing rather than hedonic pleasure alone, aligning with the human-centric governance emphasis on meaningful well-being rather than momentary satisfaction.

The simulation generated responses for 100 synthetic respondents distributed across seven professional roles according to proportions derived from legal profession surveys (percentages may not sum to 100% due to rounding): Partners (30%), Owners (27%), Marketing and Business Development professionals (17%), non-partner lawyers (13%), Trainees (9%), Managing Directors (2%), and Administrative Staff (1%). This distribution reflects the hierarchical structure of law firms while ensuring adequate representation of major stakeholder groups.

Each synthetic respondent was constructed from a demographic profile specifying age, gender, nationality, education, current location, salary range, marital status, family structure, employment level, and years of professional experience. Profile combinations were generated to represent realistic variation within each professional category. The LLM received this profile information along with the policy scenario description and typical opinions from the respondent's professional group, the latter derived from a preliminary survey of legal professionals regarding attitudes toward personal branding support.

The simulation followed a sequential response protocol designed to maintain persona consistency across the 42 questionnaire items. For each item, the LLM received three inputs: the complete demographic profile and policy context, the current questionnaire item, and the full history of previous responses for that persona. This approach mirrors human survey-taking behavior, in which respondents may reference earlier answers when responding to related questions, particularly in psychological assessments probing related dimensions of well-being.

Validation Logic and Limitations: Each response underwent immediate validation to ensure compliance with the required format (a JSON object containing an integer between 1 and 6). Invalid responses triggered a retry mechanism with slightly adjusted generation parameters, with a maximum of three attempts per item. This validation ensured that all recorded data met minimum quality standards, though it does not address whether responses are psychologically meaningful.

The simulation explores plausibility rather than predictive accuracy. We examine whether generated responses exhibit expected patterns, including differentiation across professional groups, internal consistency within personas, and distributions that fall within realistic ranges for the Ryff scale. However, demonstrating plausibility is a necessary but not sufficient condition for validity. The absence of parallel data from actual legal professionals means that we cannot assess whether the simulation accurately captures how real people would respond.

The methodological limitations are substantial and require explicit acknowledgment. LLM responses reflect patterns in training data rather than genuine psychological states. The model has learned associations between demographic characteristics and language about well-being, but it does not experience well-being and cannot report on subjective states. This fundamental asymmetry between simulation and experience constrains the inferences that can be drawn.

Additionally, the approach relies on demographic generalizations that may not capture individual variation within professional categories. A simulated partner is not a specific person, but an amalgamation of patterns associated with that professional role. Real partners differ substantially from one another in ways that demographic profiles cannot capture, including personality, life history, and individual circumstances that shape well-being in ways orthogonal to professional position.

Future research must prioritize systematic validation through parallel data collection: administering the Ryff scale to practicing legal professionals while running matched simulations of demographic profiles. A statistical comparison of response distributions, factor structures, and group differences would indicate whether AI simulations produce sufficiently accurate approximations for research or governance applications.

RESULTS

Results of Study 1: Empirical QoL Outcomes

Participants in social innovation activities showed meaningful improvements in subjective well-being over the 12-month evaluation period. Most notably, overall life satisfaction increased by 0.8 points on the 0-10 scale, a substantively meaningful increase, whereas the comparison groups remained unchanged. This finding demonstrates that subjective QoL indicators captured impacts often missed by traditional economic statistics.

More importantly, qualitative data revealed that changes were driven by stronger social support, greater self-worth, and renewed engagement with one's community. Focus group participants frequently described gaining a "renewed sense of purpose" or feeling "more in control" of their lives. These narratives corroborated the expert panel analysis, which had earlier emphasized the importance of non-material, psychosocial dimensions of social innovation.

One specific pilot in inclusive education showed marked improvements in the domains of meaning and mastery, suggesting that participants felt more capable and empowered, even without measurable income gains. A participant explained: "Before I joined, I felt like my opinions didn't matter to anyone. Now I'm part of something, and people actually listen to what I have to say." This theme of recognition and voice appeared across sites and intervention types.

The sense of purpose dimension showed particular gains in programs involving skill development and community contribution. Participants in entrepreneurship support initiatives described gaining direction and motivation that extended beyond economic outcomes. As one participant noted: "It's not just about the money. I wake up with something to work toward. That changes everything about how I feel."

Expert panels emphasized that traditional evaluations would have captured participation numbers and activity completion but missed the transformations in self-perception and social integration that participants described. One public administrator observed that the subjective well-being data "revealed the hidden value of what we're doing, the parts that don't show up in any reports we usually produce."

The results offered actionable insights for policy design. Local authorities in pilot sites used the subjective well-being data to refine project implementation. In one community, the finding that physical health was not improving alongside psychological well-being prompted integration of health components into subsequent programming. This responsive adjustment illustrates how subjective QoL data can guide iterative policy refinement in ways that output metrics alone cannot support. These findings align with literature on social innovation's capacity to enhance well-being via empowerment and social capital (Caulier-Grice et al., 2012; Edwards-Schachter et al., 2012).

Results of Study 2: Simulation Outputs

The LLM-based simulation generated complete Ryff scale responses for all 100 synthetic respondents across seven professional categories, using the Bielik model with the 42-item Ryff scale administered twice (pre- and post-policy scenario), resulting in 84 total responses per synthetic respondent. Response distributions fell within the instrument's plausible range. No synthetic respondent showed extreme response patterns (all items at ceiling or floor), and within-person variance was consistent with typical human response patterns on the Ryff scale.

Baseline Patterns by Professional Role

Professional groups exhibited differentiated response patterns consistent with theoretical expectations. Partners and owners showed the highest scores on environmental mastery and autonomy, reflecting expectations that senior professionals with organizational authority would perceive greater control over their environments. Conversely, personal growth scores were highest among trainees and junior lawyers, consistent with career stages characterized by learning and development. Senior professionals showed somewhat lower personal growth scores, potentially reflecting a plateau in professional development or a shift toward consolidating rather than expanding capabilities.

Policy Intervention Effects

The policy intervention scenario—the introduction of personal branding support—produced distinct response patterns across professional groups. Comparing simulations with and without the policy intervention revealed the following average changes across the six Ryff dimensions:

Overall dimension changes (from largest positive to largest negative): - Environmental mastery: +0.10 - Autonomy: +0.09 - Personal growth: +0.08 - Positive relations with others: +0.07 - Purpose in life: +0.01 - Self-acceptance: +0.01

Patterns by professional role:

Partners/Partnerki: Showed the most uniformly positive response to the policy. All six dimensions improved, with environmental mastery showing the largest gain (+0.08). This pattern is consistent with perceiving personal branding support as affirming professional identity and supporting strategic positioning.

Managing Directors: Showed strong positive responses across most dimensions, with positive relations showing the largest improvement (+0.32) and environmental mastery (+0.25). However, self-acceptance decreased slightly (-0.04).

Marketing/Business Development: Five dimensions improved with personal growth (+0.17) and positive relations (+0.16), showing the largest gains. Self-acceptance decreased slightly (-0.05). The elevated scores potentially reflect anticipated influence over implementation processes.

Lawyers (non-partners): Mixed pattern with four dimensions improving and two declining. Autonomy showed the largest gain (+0.08), but self-acceptance declined (-0.03).

Trainees/Aplikanci: Showed the most mixed response. Personal growth improved (+0.06), consistent with viewing the policy as a career development opportunity. However, four dimensions declined, including autonomy (-0.06) and self-acceptance (-0.10). This pattern suggests sensitivity to how policies are framed—as expectations rather than optional opportunities.

Owners/Właściciele: Most neutral response with three dimensions improving slightly and three declining slightly. The largest gains were in purpose in life (+0.02) while environmental mastery showed the largest decline (-0.04).

Administrative staff: Strong improvements in environmental mastery (+0.36), autonomy (+0.29), and self-acceptance (+0.29), but declines in positive relations (-0.07) and purpose in life (-0.07).

Limitations and Artifacts

These patterns demonstrate the simulation's capacity to generate differentiated responses reflecting plausible group interests. However, the results must be interpreted cautiously. Without validating them against actual human responses, we cannot confirm that they accurately reflect how real professionals would respond. The simulation demonstrates technical feasibility and generates hypotheses worth testing, but does not provide evidence that AI-generated well-being assessments can substitute for empirical research.

The proof-of-concept also revealed artifacts requiring attention in future development. Analysis indicated 29 positive changes, 13 negative changes, and 0 neutral changes across all dimension-by-role combinations. While the overall direction appears positive, the synthetic nature of the data means no causal conclusions can be drawn.

DISCUSSION

The combined findings support the conceptual claim that empirical QoL data and AI-based simulations can function as methodologically complementary tools for human-centric governance. Study 1 demonstrated that subjective well-being measures capture meaningful social impacts invisible to traditional metrics, while Study 2 showed that LLM-based simulation can generate differentiated well-being responses that reflect plausible group interests. Together, these studies illustrate a hybrid approach that links retrospective measurement of lived experience with anticipatory exploration of policy alternatives.

The Study 1 findings align with and extend existing research on subjective well-being in policy evaluation. The observed improvements in life satisfaction, meaning, and social connectedness mirror patterns documented in social innovation literature, where participatory programs consistently enhance psychosocial outcomes even when material circumstances remain unchanged (Edwards-Schachter et al., 2012). The observed 0.8-point increase in life satisfaction suggests a substantively meaningful change, although the exploratory design does not support formal statistical inference.

The qualitative findings illuminate mechanisms that connect intervention activities to well-being outcomes. Participants described gaining voice, recognition, and purpose through collective activity, themes that resonate with psychological theories of eudaimonic well-being, which emphasize autonomy, competence, and relatedness as fundamental human needs (Ryan & Deci, 2001). These narrative accounts provide the “mechanistic” understanding that Pawson (2013) identifies as essential for moving from retrospective evaluation to prospective policy design. Rather than simply documenting that an intervention worked, the qualitative data suggest why and for whom it worked, generating transferable insights for future programming.

Study 2's simulation results must be interpreted more cautiously given the absence of empirical validation. The differentiated response patterns across professional groups are theoretically plausible: senior professionals scoring higher on environmental mastery and autonomy, junior professionals scoring higher on personal growth, and different groups showing varying responses

to the policy intervention. However, theoretical plausibility does not establish empirical accuracy. The simulation demonstrates that LLMs can generate outputs that look like well-being data, but whether those outputs predict actual human responses remains untested.

This limitation connects to broader debates about the validity of AI-based behavioral simulation. Chen et al. (2024) showed that LLMs can replicate field experiment outcomes with reasonable accuracy, but their success varied across domains and populations. Mittelstadt et al. (2024) found that LLMs perform well on situational judgment tests, but these assessments measure reasoning about appropriate behavior rather than subjective psychological states. Well-being assessment differs from behavioral judgment in that it requires reporting on internal experience, which LLMs cannot authentically do because they lack internal experience.

The gap between language patterns about well-being and actual well-being experience represents a fundamental constraint on simulation validity. When the LLM generates a response indicating high autonomy for a simulated partner, it is drawing on patterns in training data that associate senior professional roles with language about autonomy. This may or may not correspond to how a real partner would respond to the Ryff scale item. The simulation reflects learned associations between demographic characteristics and well-being discourse, but these associations may be shaped by media representations, professional stereotypes, or other influences that diverge from actual psychological realities.

The practical implications differ substantially between the two studies. Study 1 provides direct evidence that policymakers can use to justify the use of subjective well-being assessments in routine evaluation practice. The findings support the institutionalization of QoL data collection alongside traditional metrics, consistent with OECD recommendations that subjective well-being measurement become standard practice in policy evaluation (OECD, 2023). The Impactometer demonstrates a feasible approach that captures meaningful variation while remaining practical for field implementation.

Study 2's implications are more provisional. The simulation shows technical feasibility for generating differentiated well-being responses, suggesting potential value for hypothesis generation and exploratory analysis. If validated, such tools could enable rapid prototyping of policy alternatives, allowing decision-makers to explore how different stakeholder groups might respond before committing to implementation. This aligns with Olejniczak et al.'s (2020) vision of policy design as hypothesis testing, in which interventions are treated as behavioral experiments to be refined through iterative testing.

However, the critical caveat is validation. Until systematic comparisons demonstrate that AI-generated responses approximate actual human responses with acceptable accuracy, simulation tools should supplement, rather than substitute for, traditional research methods. The appropriate use case is generating hypotheses for empirical testing, not providing definitive answers about stakeholder well-being. Decision-makers using simulation outputs without empirical grounding risk basing policy on AI-generated artifacts rather than genuine human experience, precisely the opposite of human-centric governance.

The integration of the two studies illustrates the hybrid governance model we propose. Study 1's empirical grounding informs Study 2's simulation design: the Ryff scale provides validated constructs, the policy scenario reflects realistic organizational decisions, and the demographic profiles draw on professional surveys. Conversely, Study 2's outputs generate hypotheses that could guide future empirical work. The observed sensitivity of junior professionals to how personal branding support is communicated suggests a specific question for empirical investigation: does framing such policies as opportunities versus expectations produce different well-being outcomes among trainees and early-career lawyers?

This reciprocal relationship between empirical research and simulation represents the potential contribution of the hybrid approach. Neither component alone addresses the full governance challenge. Empirical QoL research provides grounded evidence but cannot anticipate responses to untested interventions. AI simulation can explore counterfactual scenarios but lacks the authentic grounding that only human experience can provide. Combined, they offer a more complete toolkit for human-centric governance, though the combination requires careful attention to the distinct limitations of each approach.

Theoretical, Methodological, and Practical Contributions

This article makes three interconnected contributions to scholarship on human-centric governance, QoL research, and AI applications in policy and management.

The theoretical contribution lies in articulating a hybrid governance model that integrates retrospective measurement and anticipatory simulation within a coherent framework. Existing literature treats QoL assessment and AI-based behavioral simulation as separate domains, with QoL scholarship focused on measurement methodology and policy evaluation while AI governance research explores simulation capabilities and ethical constraints. We argue that effective human-centric governance requires bridging these domains by using empirical well-being data to ground simulation design and simulation outputs to inform empirical research priorities. This reciprocal relationship offers a path toward governance that is both evidence-based and forward-looking, addressing the temporal mismatch that limits current evidence-based policy frameworks.

The theoretical model responds to identified gaps in governance literature. Olejniczak et al. (2020) called for treating policy design as hypothesis testing but offered limited guidance on how to test hypotheses before full-scale implementation. Borys (2014) identified temporal mismatches between data availability and decision-making needs but did not propose mechanisms to bridge this gap. Our hybrid model addresses both concerns: empirical QoL data provides the behavioral grounding for hypotheses, while simulation enables low-cost testing of policy alternatives before implementation. The approach aligns with broader calls for anticipatory governance that integrates human-centered values into prospective decision-making (OECD, 2024; OECD, 2017). The methodological contribution involves demonstrating a structured empirical-simulation research design. Study 1 illustrates how mixed-methods QoL assessment can be integrated into policy evaluation, combining validated scales with qualitative narrative analysis to capture both the magnitude and mechanisms of changes in well-being. The Impactometer approach shows that measuring subjective well-being in field settings is feasible and yields actionable insights for policy refinement. Study 2 presents a replicable simulation pipeline that other researchers can adapt for different populations, instruments, and policy scenarios, with explicit documentation of technical parameters, prompt structures, and validation procedures.

The simulation methodology introduces specific innovations worth highlighting. The sequential response protocol, in which each questionnaire item is presented with the full history of previous responses, mirrors human survey-taking behavior and enhances persona consistency across assessments. The integration of typical group opinions derived from preliminary surveys provides empirical grounding that purely synthetic approaches lack. The use of a culturally-appropriate language model (Bielik for Polish contexts) demonstrates attention to linguistic and cultural validity that international models may not achieve. These methodological choices reflect lessons from AI simulation research and could inform future studies seeking to generate behaviorally-realistic synthetic data.

The practical contribution offers guidance to policymakers and organizational leaders on the responsible, evidence-informed use of well-being assessments and AI tools in decision-making. Study 1 demonstrates that subjective QoL data can reveal social value invisible to traditional metrics, supporting the business case for institutionalizing well-being measurement in evaluation practice. Organizations and public agencies seeking to align with human-centric governance principles can use the Impactometer approach as a template for assessing whether their interventions genuinely enhance people's lives.

For AI applications, the practical guidance emphasizes caution and complementarity. The simulation demonstrates technical feasibility but not predictive validity, underscoring that AI-generated well-being data cannot substitute for authentic human input. Appropriate use cases include generating hypotheses, exploring stakeholder diversity, and identifying potential differential impacts that warrant empirical investigation. Decision-makers should treat simulation outputs as preliminary exploration rather than definitive evidence, maintaining human oversight and empirical grounding throughout the governance process. This guidance aligns with regulatory frameworks emphasizing human-centric AI (European Commission, 2025) and ethical principles requiring human oversight of AI-assisted decisions (OECD, 2019).

Limitations and Future Research

Both studies face limitations that constrain the generalizability of findings and point toward priorities for future research.

Study 1 employed a quasi-experimental design without random assignment, introducing potential selection bias. Participants who chose to engage with social innovation activities may have differed systematically from comparison group members in motivation, social resources, or baseline well-being trajectories. While the observed improvements exceeded comparison group changes, we cannot definitively attribute these differences to the intervention alone. Future research should employ randomized designs when ethically and practically feasible, or use more sophisticated quasi-experimental methods, such as propensity score matching, to strengthen causal inference.

The sample size and geographic scope of Study 1 limit generalizability beyond the Polish social innovation context. The observed effects may depend on specific cultural, economic, or political conditions that differ across settings. Replication across diverse contexts would strengthen confidence in the transferability of both the measurement approach and the substantive findings regarding the impacts on well-being.

Study 2's limitations are more fundamental. The absence of parallel human data means that simulation accuracy cannot be assessed. The patterns observed may reflect genuine demographic associations, LLM training artifacts, or a combination of both. Without validation against actual legal professionals' responses to the Ryff scale under the specified policy scenario, we cannot determine which interpretation applies. This limitation renders the simulation a proof of concept rather than a validated research tool. The simulation's reliance on demographic generalizations may underestimate individual variation within professional categories. Real people differ in ways that demographic profiles cannot capture, including personality traits, life histories, current circumstances, and idiosyncratic factors that shape responses to well-being. The simulation's reduced within-group variance compared with typical human samples suggests that this limitation has practical consequences for the realism of the output.

The single policy scenario and professional context constrain the generalizability of simulation findings. Personal branding support for legal professionals may involve different dynamics than those in other settings. The methodology's broader applicability to diverse policy scenarios, professional contexts, and demographic populations remains untested. Future development should systematically explore these boundaries through replication across varied conditions.

Future research priorities emerge directly from these limitations. Immediate needs include systematic validation through parallel data collection, administering the Ryff scale to practicing legal professionals while running matched demographic-profile simulations. A statistical comparison of response distributions, factor structures, and group differences would indicate whether AI simulations produce sufficiently accurate approximations for research applications. Validation should examine not only overall score distributions but also subtle patterns, including response variance, inter-item correlations, and demographic effect sizes. Methodological refinement should enhance prompt optimization, demographic representation, and the sophistication of intervention modeling. The experimental approach could be extended to other psychological instruments, professional contexts, and longitudinal scenarios that model changes in well-being over time rather than single-point assessments. Such extensions would test whether the methodology generalizes beyond the initial test case.

Ethical framework development must ensure that simulated psychological data never substitute for authentic human assessment and establish oversight mechanisms for the responsible deployment of AI in governance applications. Research should address the ethical implications of "synthetic citizens" in policy consultation, including questions about representation, consent, and the potential for simulation to displace rather than complement human participation in governance.

Cross-context replication should test whether the hybrid approach functions across different policy domains, cultural settings, and organizational contexts. The combination of empirical QoL assessment and AI simulation may be more or less valuable depending on the availability of existing well-being data, the nature of the policy decisions at stake, and the institutions' capacity to act on the generated insights.

CONCLUSIONS

Human-centric governance requires tools that connect empirical well-being data with anticipatory insights about policy impacts. This article has demonstrated a hybrid approach that combines subjective QoL indicators with LLM-based simulations to address the structural gap between retrospective measurement and prospective design, which limits current evidence-based policy frameworks. Study 1 showed that subjective well-being measures capture meaningful impacts invisible to traditional metrics. Participants in social innovation activities reported meaningful improvements in life satisfaction, sense of meaning, autonomy, and social connectedness, which would have gone undetected by conventional output measures. The qualitative findings revealed mechanisms of recognition, voice, and collective contribution through which participatory activities enhanced psychological well-being. These results support the institutionalization of subjective QoL assessment in routine policy evaluation, consistent with international frameworks that emphasize well-being as a governance priority.

Study 2 demonstrated the technical feasibility of using LLMs to generate differentiated well-being responses that reflect plausible professional group interests. The simulation produced patterns consistent with theoretical expectations about how different stakeholder groups might experience organizational policy changes. However, the absence of empirical validation means that these outputs remain hypotheses rather than evidence. The appropriate use of such tools is for exploratory analysis and hypothesis generation, not for a definitive assessment of stakeholder well-being.

The combination of empirical and simulation approaches provides a more comprehensive toolkit for human-centric governance than either approach alone. Empirical QoL research provides the grounding in authentic human experience that simulation cannot achieve. Simulation enables exploration of counterfactual scenarios that empirical research cannot address due to practical constraints on time and resources. Together, they support a governance cycle in which empirical data informs simulation design, simulation outputs guide research priorities, and validated findings inform policy decisions.

Pursuing this integrated approach will require continued interdisciplinary collaboration among QoL researchers, AI scientists, policy practitioners, and governance scholars. It will require rigorous validation research that tests simulation outputs against authentic human responses across diverse contexts, and it will require ethical frameworks that maintain human centrality in governance processes, ensuring that AI tools serve rather than supplant human participation and oversight.

If developed responsibly, the potential contribution to governance quality is substantial: policies and organizational decisions that more closely align with the subjective well-being of those they affect, evaluated with sensitivity to lived experience and designed with anticipation of human responses. This vision remains aspirational, but the studies presented here suggest that the methodological foundations are achievable and worthy of continued development.

REFERENCES

- Alatartseva, E., & Barysheva, G. (2015). Well-being: Subjective and objective aspects. *Procedia - Social and Behavioral Sciences*, 166, 36–42. <https://doi.org/10.1016/j.sbspro.2014.12.479>
- Asfour, M., & Murillo, J. C. (2023). Harnessing large language models to simulate realistic human responses to social engineering attacks: A case study. *International Journal of Cybersecurity Intelligence & Cybercrime*, 6(2), 21–49. <https://doi.org/10.52306/2578-3289.1172>
- Borys, T. (2014). Wybrane problemy metodologii pomiaru nowego paradygmatu rozwoju–polskie doświadczenia. *Optimum. Studia Ekonomiczne*, 3. <https://doi.org/10.15290/ose.2014.03.69.01>
- Burchard-Dziubińska, M. (2022). Post-GDP reality: Towards a new paradigm for measuring economic welfare. *Economic and Environmental Studies*, 22(1), 5–23.
- Caulier-Grice, J., Davies, A., Patrick, R., & Norman, W. (2012). Defining social innovation: A deliverable of the project “The theoretical, empirical and policy foundations for building social innovation in Europe” (TEPSIE). European Commission, DG Research.
- Chen, Y., Hu, Y., & Lu, Y. (2024). *Simulating field experiments with large language models* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2408.09682>
- Council of the European Union. (2019, November 26). Council conclusions on the Economy of Wellbeing (2019/C 400/09). *Official Journal of the European Union*, C 400, 9–14. [https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019XG1126\(06\)&rid=5](https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019XG1126(06)&rid=5)
- Edwards-Schachter, M., Matti, C., & Alcántara, E. (2012). Fostering quality of life through social innovation: A living lab methodology study case. *Review of Policy Research*, 29(6), 672–692. <https://doi.org/10.1111/j.1541-1338.2012.00588.x>
- European Commission. (2025). *Regulatory framework for artificial intelligence*. Shaping Europe's digital future. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>
- Fedoniuk, L., & Leśniak-Moczuk, A. D. (2021). Oddziaływanie cyberświata na kondycję jednostki i społeczeństw. Nierówności Społeczne a Wzrost Gospodarczy, 68(4), 89–113. <https://doi.org/10.15584/nsawg.2021.4.5>
- Kalamucka, W. (2023). Wskaźniki społeczne jako narzędzia diagnozy i monitorowania rozwoju [Social indicators as tools for diagnosis and monitoring of development]. *Annales Universitatis Mariae Curie-Skłodowska, sectio B – Geographia, Geologia, Mineralogia et Petrographia*, 78(0), 115–130. <https://doi.org/10.17951/b.2023.78.0.115-130>
- Łabędź, A. (2023). Zarządzanie dobrostanem pracowniczym w organizacji [Managing employee well-being in the organization]. In M. Lisiński & J. Dzieńdziora (Eds.), *Paradygmaty współczesnego zarządzania z perspektywy młodych naukowców* (pp. 153–166). Akademia WSB.
- Mittelstädt, J. M., Maier, J., Goerke, P., Zinn, F., & Hermes, M. (2024). Large language models can outperform humans in social situational judgments. *Scientific Reports*, 14, Article 27449. <https://doi.org/10.1038/s41598-024-79048-0>
- OECD. (2017). *Behavioural insights and public policy: Lessons from around the world*. OECD Publishing. <https://doi.org/10.1787/9789264270480-en>
- OECD. (2019). *OECD Principles on Artificial Intelligence* (Legal Instrument No. OECD/LEGAL/0449). <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- OECD. (2023). *Subjective well-being measurement: Current practice and new frontiers* (WISE Working Paper No. 2023/5). OECD Publishing. <https://doi.org/10.1787/2226583X>
- OECD. (2024). Framework for Anticipatory Governance of Emerging Technologies. OECD Publishing. https://www.oecd.org/en/publications/2024/04/framework-for-anticipatory-governance-of-emerging-technologies_14bf0402.html
- Olejniczak, K., Śliwowski, P., & Leeuw, F. (2020). *Comparing Behavioral Assumptions of Policy Tools: Framework for Policy Designers*. Journal of Comparative Policy Analysis: Research and Practice, 22(6), 498–520. <https://doi.org/10.1080/13876988.2020.1808465>
- Pawson, R. (2013). *The science of evaluation: A realist manifesto*. SAGE Publications.
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology*, 52(1), 141–166. <https://doi.org/10.1146/annurev.psych.52.1.141>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069–1081. <https://doi.org/10.1037/0022-3514.57.6.1069>

FUNDED RESEARCH PROJECTS

- Wojtkiewicz, K. B. (2021). An integrated model of subjective quality of life indicators for evaluating social innovation policies (Research Package 9). Grant No. EOG/21/K4/W/0044W/0167, EEA Grants 2014–2021.
- Wojtkiewicz, K. B. (2024). CAM4QoL: Complex adaptive management—Leveraging Nash equilibrium insights for QoL-oriented governance. Grant No. 51/2024/FRBN/G, SWPS University Research Development Program.