Synthetic Personas: Enhancing Demographic Response Simulation Through Large Language Models and Genetic Algorithms

Morten Grundetjern, Per Arne Andersen, and Morten Goodwin

University of Agder, Grimstad, Norway

Abstract. Understanding diverse demographic groups presents a significant challenge in market research. In this paper, we introduce a novel system that integrates large language models with genetic algorithms to create synthetic personas capable of generating feedback that approximates real-world human responses. Our experimental evaluation demonstrates that synthetic personas not only exhibit age-differentiated technology usage patterns consistent with documented trends but also benefit from genetic algorithm optimization, which improves response accuracy from 60.4% to 78.5% on training questions and from 62.6% to 68.8% on hidden questions—outperforming human estimators. Moreover, the optimized personas achieve a 51.1% better correspondence with actual income distributions compared to random profiles. This approach makes it possible to rapidly generate feedback without requiring participants, facilitates iterative follow-ups, and systematically enhances demographic representativeness

Keywords: Synthetic Personas Large Language Models Genetic Algorithms Demographic Modeling Survey Response Simulation

1 Introduction

The accurate modeling of diverse demographic groups represents a significant challenge in market research and product development [1]. While traditional methodologies such as focus groups and surveys provide valuable insights, they often lack the adaptability and scale required to capture rapidly evolving user preferences [2]. Recent advances in Large Language Models (LLMs) offer promising capabilities for addressing these methodological limitations, yet their application in emulating specific demographic feedback patterns remains underexplored.

This research investigates the efficacy of combining LLMs with Genetic Algorithms (GAs) to create synthetic personas capable of generating feedback that closely approximates real-world human responses. We define synthetic personas as instances of LLMs conditioned with textual descriptions designed to represent specific demographic attributes, personality traits, interests, and behavioral tendencies of hypothetical individuals. To the best of our knowledge, no prior research has explored the systematic optimization of LLM-based synthetic personas using evolutionary algorithms to enhance demographic response simulation, nor has previous work quantitatively evaluated the capacity of such systems to generate responses that align with established demographic patterns across both seen and unseen questions.

Our methodology introduces three primary contributions. First, we develop a novel approach that integrates LLMs with GAs to optimize synthetic persona selection, ensuring responses align more closely with target demographic patterns. Second, we demonstrate how this system enables iterative dialogue capabilities that traditional methods cannot efficiently support—while conventional surveys require developing new questionnaires to explore emerging topics, our system facilitates immediate follow-up questions based on initial responses. Third, we provide quantitative evidence of demographic consistency in synthetic persona responses.

DOI:10.5121/ijci.2025.140202

Experimental results demonstrate that our synthetic personas exhibit behavior patterns that align with established demographic norms, particularly regarding age-related response variations. Through iterative fine-tuning using GAs, we achieve substantial improvements in persona generation quality and response accuracy. Our findings indicate that optimized personas demonstrate increasing proficiency in answering both previously encountered questions and novel inquiries, suggesting robust generalization capabilities. Further analysis reveals the importance of sequential context in question presentation and its effects on response patterns. Notably, the income distribution patterns in our optimized "highest-scoring persona" group showed close correspondence with real-world demographic data, suggesting broader applications beyond our primary research objectives.

By exploring this intersection of artificial intelligence and market research, we provide insights that can shape future strategies in understanding and engaging with diverse user demographics.

2 Literature Review

This section explores the evolution of natural language processing (NLP) techniques, tracing their development from early statistical methods to advanced transformer-based architectures that enable the synthetic persona approach proposed in our research.

2.1 Evolution of NLP Techniques

The ability to emulate diverse demographic perspectives has become increasingly viable with advances in text generation models. These models have undergone significant transformations, evolving from basic statistical methods like Naive Bayes classifiers [3] to sophisticated neural architectures such as BERT and Generative Pre-trained Transformers [4,5].

Early text generation relied primarily on Markov chain modeling, pioneered by Andrei A. Markov through his analysis of letter sequences in literature. Claude E. Shannon extended this approach by introducing a mathematical theory of communication that employed Markov chains to generate text with structural similarities to natural language [6]. While groundbreaking, these early attempts produced output like "IN NO IST LAT WHEY CRATICT FOURE BIRS GROCID"—text that exhibited linguistic patterns but lacked semantic coherence.

Despite their historical significance, these statistical methods demonstrated fundamental limitations in capturing contextual relationships and generating coherent text. The field subsequently shifted toward deeper semantic understanding, with Hidden Markov Models (HMMs) representing an important advancement, particularly in speech recognition applications [6]. However, these approaches still failed to model long-term dependencies and contextual nuances necessary for creating realistic synthetic personas capable of providing authentic demographic feedback.

2.2 Neural Network Architectures for NLP

As computational resources expanded in the early 2010s, Recurrent Neural Networks (RNNs) emerged as a dominant paradigm for sequential data processing [7]. Unlike feed-forward networks, RNNs maintain internal states that function as memory for previous inputs, making them particularly suitable for processing text data where meaning depends on sequential context. This capacity to preserve contextual information represented a significant advancement toward building synthetic personas with consistent response patterns.

Despite their theoretical advantages, RNNs faced practical challenges related to training stability, particularly the vanishing and exploding gradient problems [8]. Long Short-Term Memory (LSTM) networks, introduced by Hochreiter and Schmidhuber, addressed these limitations through specialized gate mechanisms that regulated information flow through the network [9]. These innovations enabled more effective learning of long-range dependencies in text.

The sequence-to-sequence (seq2seq) architecture further extended these capabilities by implementing an encoder-decoder structure [10]. This design compressed input sequences into fixed-size context vectors before generating variable-length outputs, enabling applications like machine translation. While these models demonstrated impressive capabilities, they still exhibited significant limitations in maintaining coherent long-form responses and incorporating broader world knowledge—critical requirements for convincing demographic emulation. The need for richer contextual representation and more efficient training mechanisms remained unmet until the advent of attention-based models.

2.3 Transformer Architecture and Large Language Models

The introduction of the Transformer architecture by Vaswani et al. marked a paradigm shift in NLP [11]. By replacing recurrent connections with multi-head attention mechanisms, Transformers enabled parallel processing of entire sequences while maintaining awareness of positional relationships. This innovation dramatically improved both training efficiency and model performance across diverse language tasks.

BERT (Bidirectional Encoder Representations from Transformers) further revolutionized the field by introducing deep bidirectional pre-training that conditioned on both left and right contexts simultaneously [4]. This approach created contextually rich representations that could be fine-tuned for specific downstream tasks with minimal architectural modifications. The bidirectional nature of BERT represented a significant advance in language understanding capabilities.

Subsequent developments in generative models, particularly GPT-3 and its successors, expanded text generation capabilities to include increasingly sophisticated applications such as interactive dialogue, creative writing, and code generation [12,13]. These advancements enable more realistic emulation of human-like responses across diverse demographic profiles—a critical foundation for our synthetic persona methodology. The state-of-the-art models available at the time of writing, including GPT-4 and LLaMA-2, establish new benchmarks in generating contextually appropriate, nuanced responses that can effectively represent specific demographic perspectives [14, 15].

2.4 Prompting Strategies and Techniques

The effective utilization of large language models depends significantly on prompting strategies—the methods used to condition model outputs toward desired behaviors or response patterns. As these models grow in size and capability, prompting has emerged as a crucial research area for enhancing performance without requiring additional training [**?**].

Chain of Thought Methods Kojima et al. demonstrated that LLMs possess latent reasoning capabilities that can be activated through appropriate prompting strategies [16]. Their research introduced Zero-shot Chain of Thought (CoT) prompting, showing that simple instructional phrases like "Let's think step by step" significantly enhance model performance on complex reasoning tasks without requiring exemplars. This finding suggests that LLMs encode sophisticated cognitive processes that can be elicited through properly structured prompts—a capability we leverage in our synthetic persona design.

Building on this foundation, Chung et al. explored instruction fine-tuning approaches that explicitly train models to produce step-by-step reasoning paths [17]. Their work demonstrated that scaling both model size and task diversity yields compounding benefits, particularly when combined with chain-of-thought training procedures. This approach enhances models' ability to provide structured, logical responses aligned with demographic characteristics.

Fu et al. further refined prompting techniques by introducing complexity-based prompting [18]. This approach systematically varies the complexity of reasoning chains in prompts, demonstrating that more elaborate prompts consistently yield improved performance on challenging tasks. Their methodology produced substantial accuracy improvements across benchmark datasets, with gains ranging from 5.3 to 18 percentage points on state-of-the-art models. The demonstrated relationship between prompt complexity and response quality informs our approach to constructing demographic-specific prompt templates.

Self-Consistency Methods Wang et al. introduced self-consistency as a complementary approach to chain-of-thought prompting [19]. Rather than relying on a single reasoning path, this method generates multiple potential solutions and selects the most consistent answer. The underlying principle—that complex problems often have multiple valid reasoning approaches leading to the same correct solution—aligns with our objective of producing stable, representative demographic responses across varying contexts. Their empirical results demonstrated significant performance improvements without requiring model retraining, establishing self-consistency as a computationally efficient enhancement to standard prompting techniques.

These prompting innovations collectively enable more sophisticated interactions with language models, allowing researchers to elicit specific response patterns aligned with demographic characteristics. By combining structured prompts with appropriate demographic constraints, we can guide model outputs to reflect authentic variations in perspective across different population segments.

2.5 Traditional Market Research Methods

Prior to the emergence of computational approaches to demographic modeling, market research relied primarily on direct data collection methods with inherent limitations in scale, cost, and adaptability. Understanding these traditional approaches provides important context for appreciating the advances offered by synthetic persona methodologies.

Surveys and questionnaires have historically served as primary instruments for gathering demographic insights, offering structured frameworks for data collection [20]. However, these methods involve considerable resource investments in design, distribution, and analysis. Their effectiveness depends heavily on participant engagement, with response rates averaging between 5-30% depending on methodology [21]. Additionally, survey design introduces potential biases through question framing and sampling methods that may limit demographic representation.

Focus groups and in-depth interviews provide qualitative depth through interactive engagement with population samples [22]. While these approaches yield rich contextual insights, they face significant constraints in terms of geographic coverage, cost scalability, and potential moderator bias. The resource-intensive nature of these methods typically restricts sample sizes, limiting statistical power and generalizability.

Traditional market research methods also suffer from temporal limitations, with typical studies requiring weeks or months to progress from design to analysis. This latency creates challenges for iterative research, as follow-up questions or hypothesis refinement necessitate initiating new data collection cycles rather than building directly on previous interactions [2]. The static nature of these approaches contrasts sharply with the dynamic, adaptive capabilities enabled by synthetic persona methodologies.

These methodological constraints have motivated the exploration of computational approaches that complement traditional techniques. By combining the strengths of large language models with structured optimization through genetic algorithms, our synthetic persona approach addresses fundamental limitations in traditional market research while maintaining the authentic demographic variability that these methods aim to capture.

3 Methodology

This section details our approach to generating demographic-specific synthetic feedback using LLMs and GAs, covering system architecture, persona construction, GA optimization, and evaluation procedures.

3.1 System Architecture Overview

Our methodology integrates LLMs with evolutionary optimization to create synthetic personas that closely approximate responses from specific demographic groups. Figure 1 illustrates the workflow: (1) synthetic persona generation, (2) response collection, (3) genetic algorithm optimization, and (4) evaluation against hidden test questions.

The process generates diverse synthetic personas with defined attributes, collects their responses to survey questions, applies GA optimization to identify optimal persona configurations, and evaluates their performance on unseen questions.

3.2 Synthetic Persona Construction

We define synthetic personas as LLM instances conditioned with descriptions representing demographic attributes, personality traits, and behavioral tendencies of hypothetical individuals. Each persona includes:

- **Demographic Attributes:** Age, gender, location, education, occupation, income
- **Personality Traits:** OCEAN model parameters [23]
- **Preference Patterns:** Hobbies, media consumption, lifestyle characteristics

Our implementation emulates Norwegian demographic distributions through structured profile generation. Gendered names follows national distribution patterns, influencing name selection from Norwegian databases. Age values range from 18-70 years, while geographic distribution reflects actual population density across Norwegian counties.

Personality traits are constrained to moderate ranges (40-60 on a 100-point scale) to avoid unrealistic response patterns. When generating responses, each persona receives specialized conditioning through prompt engineering that employs Chain of Thought mechanisms [16] to enhance demographic consistency.

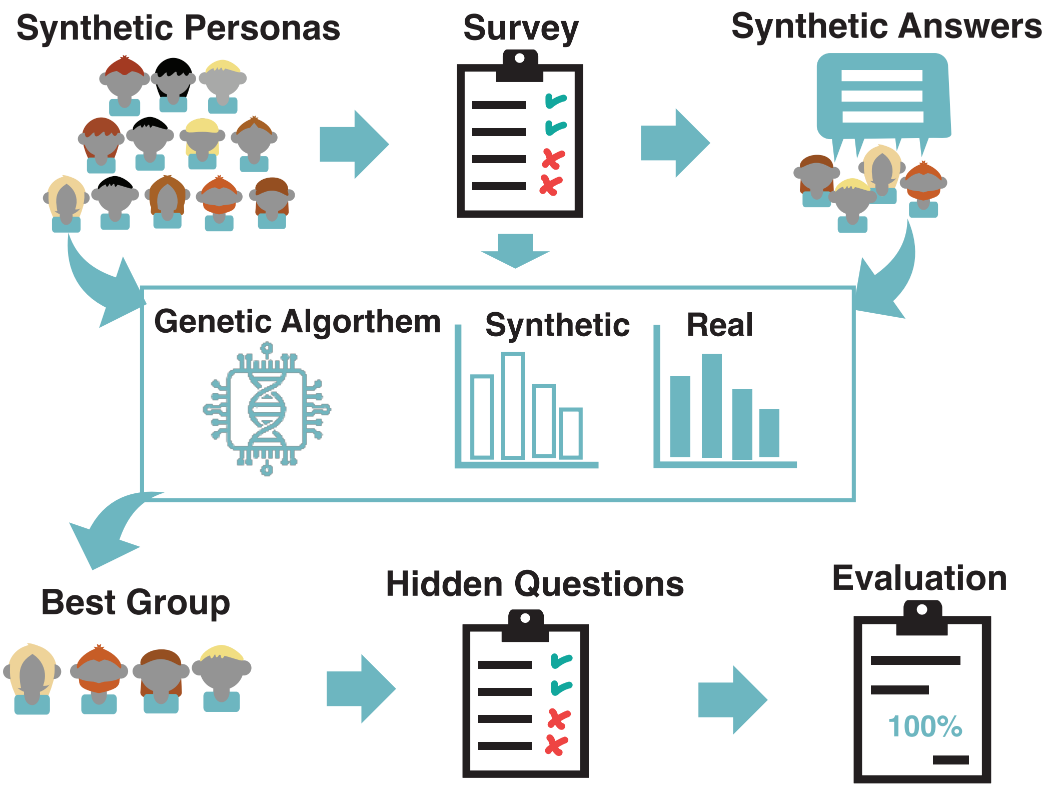

Fig. 1. System workflow showing the integration of synthetic personas with genetic algorithm optimization, from generating diverse personas to evaluating generalization performance.

3.3 Genetic Algorithm Optimization Framework

Genetic algorithms are particularly well-suited for this optimization problem compared to alternative approaches. Unlike gradient-based methods that require differentiable functions, GAs effectively navigate our discrete search space without getting trapped in local optima. Reinforcement learning and simulated annealing lack the parallel exploration that makes GAs efficient for identifying complementary persona groupings. Our approach excels in maintaining diversity and handling combinatorial complexity—though at the cost of computational requirements. This evolutionary method better models the complex relationships between demographic attributes and response patterns than deterministic optimization techniques.

We apply genetic algorithms to identify optimal subsets of synthetic personas that collectively generate feedback patterns aligned with actual demographic response distributions. In our framework, individual personas function as genes, persona groups operate as chromosomes, and multiple groups form a supergroup (population).

The optimization process begins with 50 candidate chromosomes, each representing a different configuration of synthetic personas. Our fitness function evaluates chromosomes based on:

$$\mathrm{Fitness}(C) = \alpha \cdot \mathrm{DistributionSimilarity}(C, R) + \beta \cdot \mathrm{ResponseDiversity}(C) \tag{1}$$

where $\mathrm{DistributionSimilarity}(C, R)$ measures how closely the aggregate response distribution from candidate personas C matches the reference distribution R from real demographic data, calculated as $1 - \frac{1}{n}\sum_{i=1}^{n} |C_i - R_i|$ (normalized Manhattan distance), with n

representing the number of response options across all questions; and ResponseDiversity(C) quantifies the variation in responses within the candidate group, ensuring heterogeneity in perspective, computed as the average pairwise Jaccard distance between persona response sets. Higher values for both metrics contribute to higher overall fitness.

where C represents a candidate chromosome, R represents reference demographic responses, and α and β are weighting parameters (0.8 and 0.2, respectively) determined through manual hyperparameter tuning.

We employ tournament selection ($k = 3$), single-point crossover (rate: 0.8), and mutation through random persona substitution (rate: 0.2). These genetic algorithm parameters were established through preliminary experiments that balanced exploration and exploitation within our computational constraints. The process continues for 100-120 generations or until convergence (fitness improvement < 0.001 over 5 consecutive generations), with elitism preserving the top 2 chromosomes between generations to maintain solution quality across iterations.

3.4 Evaluation Methodology

We implement a structured evaluation approach that measures both training performance and generalization capabilities. Questions are randomly partitioned into training (75%) and hidden test (25%) subsets to get sufficient statistics, with optimization performed exclusively on the training subset.

For each evaluation cycle, we measure three key metrics: distribution similarity between synthetic and actual responses, response diversity across the persona group, and the performance gap between training and test questions. Multiple evaluation cycles with different random partitions ensure robust assessment of generalization capabilities.

This methodology provides a systematic framework for generating representative demographic feedback that closely approximates human response patterns, leveraging LLMs' generative capabilities while employing evolutionary algorithms to enhance demographic alignment and response quality.

4 Results

This section presents the empirical findings from three experiments designed to evaluate our synthetic persona approach. First, we examine the demographic consistency of synthetic personas by analyzing age-based response patterns. Second, we assess the effectiveness of genetic algorithm optimization in improving response accuracy. Finally, we evaluate the alignment between optimized synthetic personas and real-world demographic distributions.

4.1 Human Baseline Experiment

To provide a realistic performance benchmark, we conducted a human evaluation experiment. Eight participants were asked to estimate response distributions using an interactive web interface. The interface presented survey questions alongside sliders corresponding to common answer options, and participants allocated percentages across these options such that the total summed to 100%. Each participant responded to approximately 20 questions, yielding 166 valid distribution.

Performance was assessed by computing the weighted mean error between the human-estimated distributions and the ground truth, using the L1 distance normalized by the maximum possible error (200 percentage points). The resulting human baseline accuracy was 63.5%, which establishes a realistic target for our synthetic methods.

4.2 Experiment 1: Demographic Consistency in Synthetic Personas

Experimental Design This experiment examines whether synthetic personas exhibit response patterns consistent with their assigned demographic attributes, focusing specifically on the relationship between age and online behavior. We investigated two research questions:

1. Can synthetic personas generate responses that reflect meaningful demographic variation based on their assigned characteristics?
2. Does modifying a specific demographic variable (age) produce predictable changes in response patterns that align with established behavioral tendencies?

We generated 200 synthetic personas distributed equally across four age brackets: 18-29 years, 30-49 years, 50-64 years, and 65+ years (50 personas per bracket). All other demographic attributes were randomly assigned according to realistic distributions. Each persona responded to the question: "Are you always online?" with one of three possible responses: "Several times a day," "Several times a week," or "Less often."

Results Figure 2 presents the distribution of responses across age brackets. The results demonstrate a clear inverse relationship between age and reported frequency of online activity. Synthetic personas in younger age brackets (18-29 and 30-49) overwhelmingly reported being online "Several times a day" (100% in both groups). This frequency progressively declined in older age brackets, with only 82.6% of personas aged 50-64 and just 7.0% of personas aged 65+ reporting daily online activity.

The most pronounced shift occurred in the 65+ demographic, where the majority (90.7%) reported being online "Several times a week" rather than daily. Additionally, a small percentage (2.3%) of this oldest group selected "Less often," a response entirely absent in the younger age brackets.

Analysis The observed response pattern closely aligns with documented age-related technology usage trends in empirical research. Selwyn et al. [24] found that older adults integrate information and communication technologies into their daily routines less frequently than younger populations. Our results quantitatively mirror this pattern, with the percentage of personas reporting daily online activity decreasing from 100% in the 18-49 brackets to 7.0% in the 65+ bracket.

This experiment provides strong evidence that synthetic personas can generate demographically consistent responses when differentiated by a single variable (age) while controlling for other factors. The clear monotonic relationship between age and response distribution demonstrates that LLMs can effectively model demographic-specific behavioral patterns when properly conditioned with relevant attributes.

4.3 Experiment 2: Genetic Algorithm Optimization Effectiveness

Experimental Design This experiment evaluates the effectiveness of genetic algorithm optimization in enhancing the alignment between synthetic persona responses and actual human survey responses. We implemented a cross-validation approach by partitioning a 20-question survey dataset into training (75%, 15 questions) and testing (25%, 5 questions) subsets. The GA optimization process operated exclusively on the training subset, while the test subset provided an unbiased assessment of generalization performance.

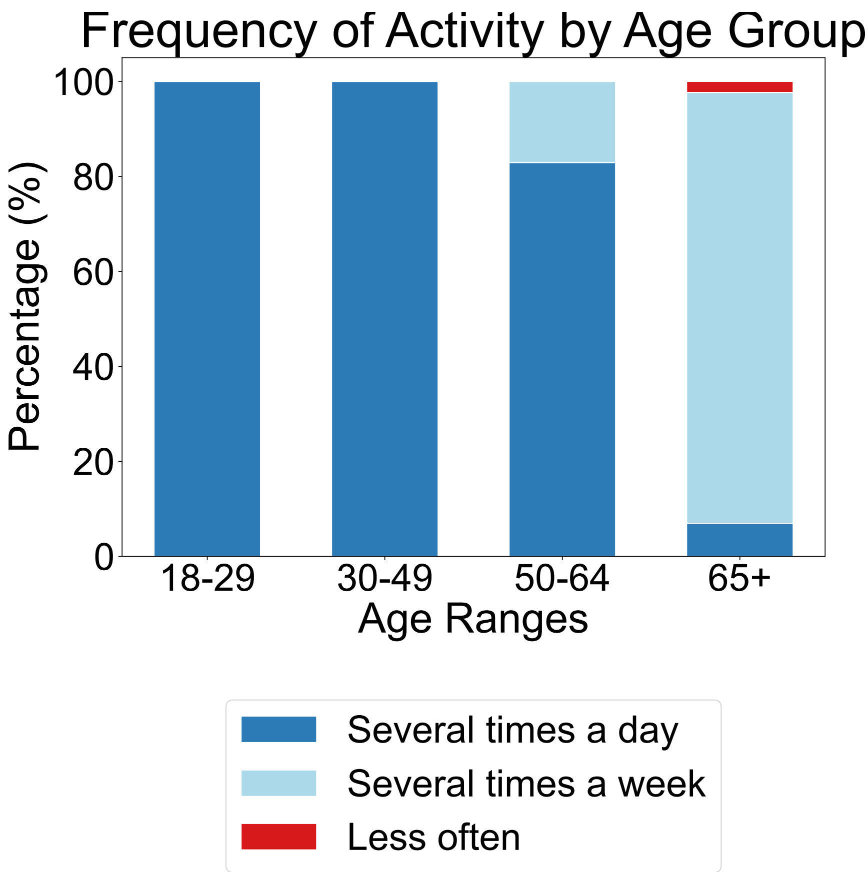

Fig. 2. Age-based variation in reported online activity frequency among synthetic personas. The graph demonstrates a systematic decline in daily online presence with increasing age, particularly pronounced in the 65+ demographic. This pattern aligns with established age-related technology usage patterns documented in prior research.

To account for potential ordering effects in sequential questions, we conducted two separate evaluation protocols: (1) random question partitioning, where testing questions were randomly selected across the full set, and (2) sequential partitioning, where the final five questions were consistently assigned to the test set. Each protocol was repeated across 10 independent trials with different random seeds to ensure statistical robustness.

For each iteration, synthetic personas responded to questions sequentially, with contextual awareness of previous questions and responses to maintain consistent answer patterns. The genetic algorithm selected optimal persona subsets by evaluating the similarity between aggregated synthetic responses and actual human response distributions on the training questions.

Results Figure 3 illustrates the progression of response accuracy across GA generations for both training and test question sets. The optimization process demonstrated substantial improvement in training set accuracy, increasing from an initial mean value of 60.4% (SD=2.3%) to 78.5% (SD=1.7%) after convergence. More importantly, test set accuracy showed corresponding improvement from 62.6% (SD=3.5%) to 68.8% (SD=2.9%), indicating effective generalization to unseen questions.

To quantify the statistical significance of these improvements, we calculated Pearson correlation coefficients between generation number and accuracy metrics. For training accuracy, we observed $r = 0.939$ with $p < 6.37 \times 10^{-75}$, indicating a strong positive relationship. Test accuracy showed $r = 0.871$ with $p < 1.03 \times 10^{-50}$, confirming that the improvement pattern extends reliably to unseen questions.

The sequential partitioning protocol (Figure 4) produced comparable results, with final training accuracy of 77.6% (SD=1.9%) and test accuracy of 67.2% (SD=3.1%). The consistent improvement pattern across both protocols demonstrates that the optimization benefits are robust to different question sampling methodologies.

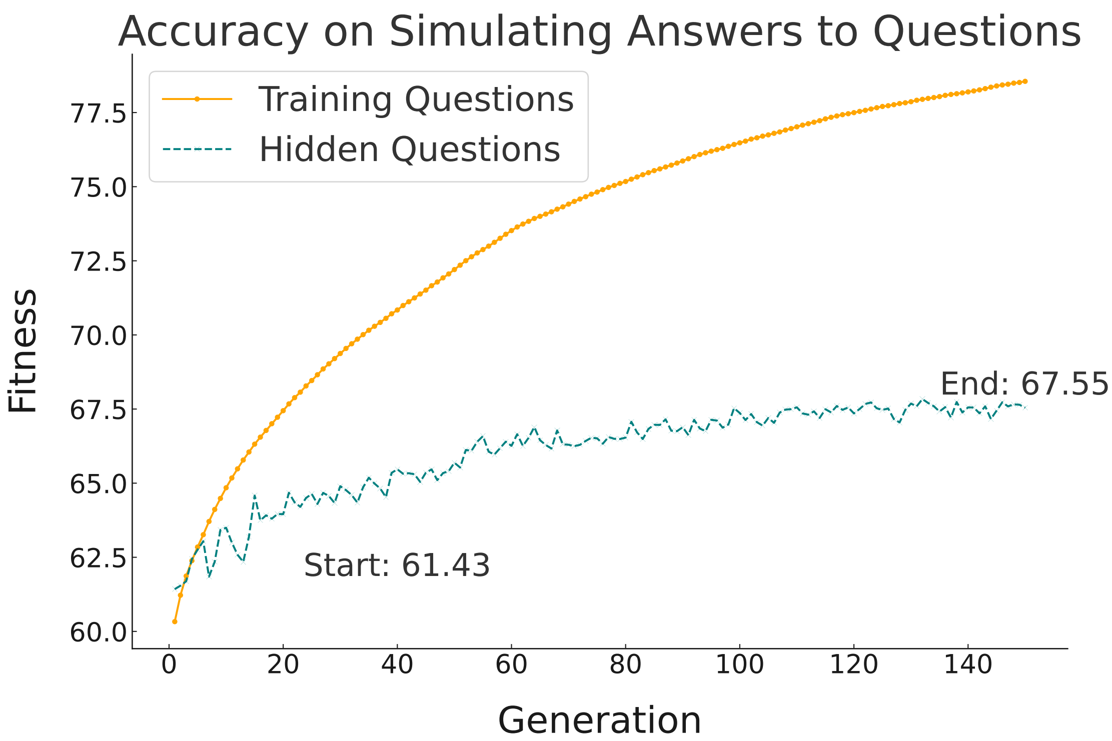

Fig. 3. Accuracy progression across genetic algorithm generations for both training questions (solid line) and unseen test questions (dashed line). The graph demonstrates consistent improvement in both metrics, with training accuracy increasing from 60.4% to 78.5% and test accuracy improving from 62.6% to 68.8%, indicating effective generalization capability.

Analysis The experimental results provide compelling evidence that genetic algorithm optimization significantly enhances the quality of synthetic persona responses. The substantial improvement in training accuracy (30.0% relative increase) demonstrates the effectiveness of the selection and evolution process in identifying optimal persona configurations. More critically, the corresponding improvement in test accuracy (10.0% relative increase) confirms that the benefits extend beyond the optimization dataset to novel questions.

The high correlation coefficients with extremely low p-values establish the statistical significance of these improvements, ruling out the possibility that the observed patterns result from random variation. Furthermore, the consistent improvement patterns across

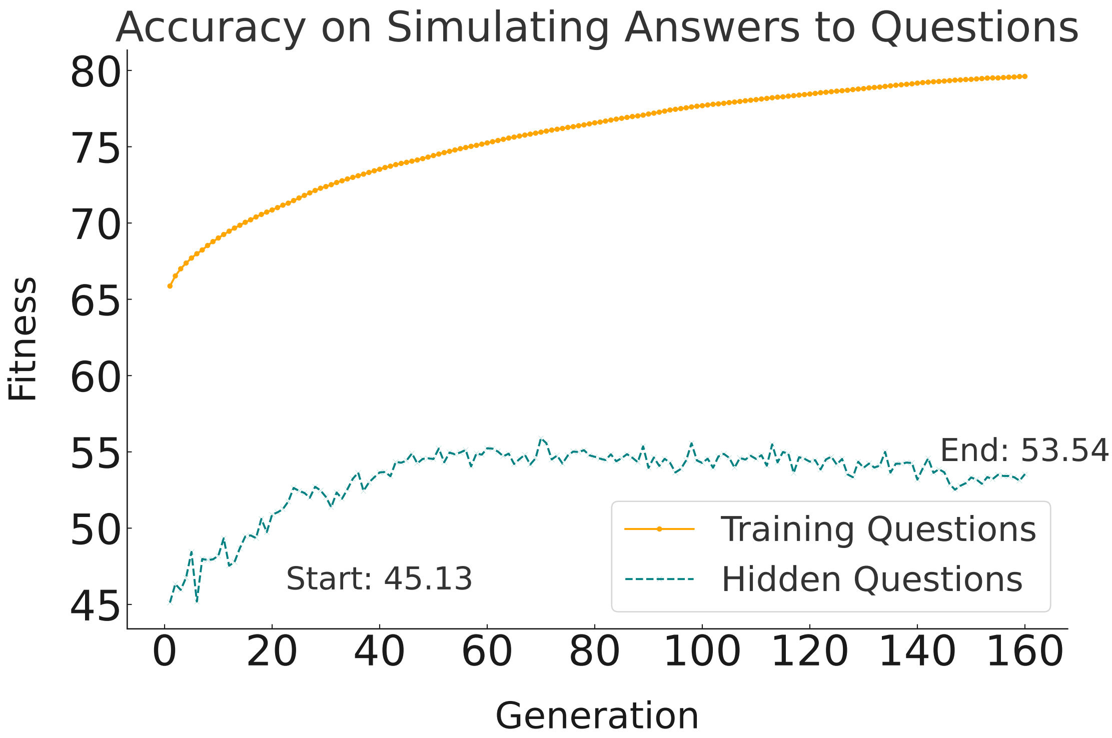

Fig. 4. Accuracy progression when specifically using the last 5 sequential questions as the test set. This alternative partitioning approach demonstrates similar improvement patterns to random partitioning, confirming the robustness of the optimization methodology against potential order effects.

different question partitioning methodologies indicate that the optimization approach is robust against potential ordering effects in sequential survey questions.

These findings confirm that synthetic personas can be systematically refined through evolutionary optimization to generate more accurate representations of target demographic response patterns. The ability to generalize to unseen questions suggests that the optimization process identifies fundamental demographic response characteristics rather than merely overfitting to specific question formulations.

4.4 Experiment 3: Demographic Distribution Alignment

Experimental Design This experiment examines the degree to which optimized synthetic personas exhibit attribute distributions that align with actual demographic data, independent of the optimization objective. We focused specifically on income distribution as a representative demographic attribute that strongly correlates with numerous social and behavioral patterns.

We compared the income bracket distribution of three groups: (1) randomly generated synthetic personas without optimization, (2) the "Highest-scoring Persona" group selected through GA optimization based on response accuracy, and (3) actual Norwegian population data from Statistics Norway (SSB) for 2017 [25]. Importantly, income distribution was not explicitly included in the optimization fitness function, allowing this analysis to assess whether response pattern optimization indirectly improves demographic representativeness.

Results Table 1 presents the comparative income distributions across the three groups. The GA-optimized "Highest-scoring Persona" group demonstrated substantially closer alignment with real-world income distribution than the unoptimized random profiles. The mean absolute deviation from real-world percentages decreased from 9.0 percentage points in random profiles to 4.4 percentage points in the optimized group, representing a 51.1% improvement in distribution accuracy.

This alignment improvement was particularly pronounced in the highest income bracket (900,000+ NOK), where random profiles showed substantial overrepresentation (28.0% vs. 7.0% in real data), while the optimized group reduced this deviation by 50% (14.0%). Similarly, the lowest income bracket (under 300,000 NOK) showed improved alignment, with the optimized group (30.0%) more closely approximating the actual distribution (35.6%) compared to random profiles (20.0%).

Table 1. Income Distribution Comparison Between Synthetic Personas and Actual Norwegian Population Data. Data collected from Statistisk sentralbyrå (SSB), Norway.

Income Category	**Highest-scoring Persona**	**Random Profiles**	**Real-world 2017**
Under 300,000 NOK	30.00%	20.00%	35.63%
300,000 - 499,999 NOK	30.00%	24.00%	31.56%
500,000 - 699,999 NOK	16.00%	16.00%	18.80%
700,000 - 899,999 NOK	10.00%	12.00%	6.97%
900,000 NOK or more	14.00%	28.00%	7.04%

Analysis The improved alignment between optimized synthetic personas and actual demographic distributions represents an important secondary finding of our research. Despite not explicitly incorporating income distribution in the optimization fitness function, the GA selection process indirectly identified persona configurations with more realistic demographic attributes.

This emergent demographic alignment suggests that response patterns and demographic attributes share underlying correlational structures that the optimization process effectively captures. By selecting personas based on response accuracy, the GA appears to indirectly favor personas with more representative demographic profiles. This finding has significant implications for the ecological validity of synthetic persona research, suggesting that optimized personas not only generate more accurate responses but also better represent the actual composition of target populations.

The relationship between response optimization and demographic alignment also provides preliminary evidence that our approach might support bidirectional inference—not only generating demographically appropriate responses from defined personas but potentially inferring likely demographic attributes from response patterns. While this aspect requires further investigation, the observed income distribution alignment suggests promising applications in demographic modeling and market segmentation.

5 Discussion

This section analyzes the implications of our experimental findings, identifies methodological limitations, and discusses broader applications of synthetic personas for demographic modeling. We examine observed behavioral patterns in large language models, analyze emergent demographic alignments resulting from genetic algorithm optimization, and consider both practical applications and ethical considerations of this approach.

5.1 Interpretation of Key Findings

Order Effects in Language Model Responses Our experiments revealed significant sensitivity of LLMs to prompt structure, particularly the presentation order of response options. When presented with multiple-choice questions, synthetic personas demonstrated a consistent bias toward options presented first in the sequence. This phenomenon parallels the well-documented ”primacy effect” in human cognitive psychology, where items appearing earlier in a sequence receive disproportionate attention and recall [26].

This observation has important methodological implications for synthetic persona research. Researchers employing LLMs for survey response simulation must carefully control for potential order effects by implementing techniques such as randomized option ordering, balanced experimental designs, or statistical corrections for position bias. Without such controls, synthetic response distributions may reflect artifacts of prompt construction rather than genuine demographic variations.

The susceptibility of LLMs to order effects also suggests fundamental similarities between artificial and human information processing mechanisms. Both appear to allocate differential attention to sequential information, raising intriguing questions about the cognitive architectures that emerge in large-scale neural language models. Further research comparing human and LLM order effects across diverse question types could yield valuable insights into both artificial and human cognition.

Emergent Demographic Alignment Through Optimization A particularly significant finding from our third experiment was the progressive alignment of synthetic persona attributes with real-world demographic distributions through genetic algorithm optimization. Despite not explicitly incorporating income distribution in our fitness function, the optimized ”Fittest Persona” group demonstrated substantially improved correspondence with actual Norwegian income patterns compared to random profiles.

This emergent alignment suggests that response patterns and demographic attributes share underlying correlational structures that were implicitly captured through our optimization process. By selecting personas based solely on response accuracy, the genetic algorithm indirectly favored configurations with more realistic demographic profiles. This observation supports the hypothesis that certain demographic characteristics correlate with distinct response patterns across diverse question types.

The evolution of more realistic income distributions without explicit optimization for this attribute demonstrates the potential of our approach to capture complex relationships between demographic factors and response behaviors. This finding has significant implications for the ecological validity of synthetic persona methodologies, suggesting that optimized personas not only generate more accurate responses but also better represent the actual composition of target populations.

5.2 Theoretical and Practical Implications

Demographic Inference Capabilities The observed convergence toward realistic demographic distributions raises intriguing possibilities for bidirectional inference between response patterns and demographic profiles. If synthetic personas optimized for response accuracy naturally converge toward realistic demographic distributions, it suggests the potential to reverse this relationship—inferring likely demographic characteristics from observed response patterns.

This capability could enable several innovative applications:

- **Enhanced Market Segmentation:** Organizations could identify likely demographic profiles of consumer groups based on response patterns to product feedback surveys, enabling more targeted product development and marketing strategies.
- **Population Modeling:** Researchers could generate demographic projections from longitudinal survey data, potentially identifying emerging demographic trends before they become apparent in traditional census data.
- **Personalized Content Delivery:** Systems could dynamically adjust content presentation based on inferred demographic characteristics without requiring explicit demographic data collection.

However, these capabilities also raise significant ethical considerations regarding privacy and consent. The ability to infer detailed demographic information from seemingly unrelated response patterns could potentially circumvent explicit data collection restrictions. This underscores the need for transparent disclosure and appropriate privacy safeguards when implementing synthetic persona technologies in practical applications. One potential method could be utilizing NEF, as described in Grundetjern et al. [27].

Implications for Survey Methodology Our findings have substantial implications for traditional survey research methodologies. The demonstrated ability of synthetic personas to generate demographically consistent responses across both seen and unseen questions suggests potential applications in survey design, testing, and analysis:

- **Survey Pre-testing:** Researchers could employ synthetic personas to identify potential biases, ambiguities, or demographic sensitivities in survey instruments before field deployment.
- **Sample Augmentation:** In cases where specific demographic segments are underrepresented in survey samples, calibrated synthetic personas could potentially augment collected data to improve representativeness.
- **Exploratory Analysis:** Researchers could use synthetic personas to conduct preliminary explorations of research questions, generating hypotheses for subsequent validation with human participants.

While our current implementation focused on technology usage patterns, this approach shows promise for broader application across various domains including consumer behavior research, political polling, healthcare preference analysis, and product development feedback. The framework's flexibility allows it to be adapted to different question domains with domain-specific optimization criteria.

Importantly, we view these applications as complementary to, rather than replacements for, traditional human participant research. The value of synthetic personas lies in their ability to enhance efficiency, expand exploratory capabilities, and identify potential methodological issues before human data collection.

5.3 Limitations and Future Directions

Methodological Limitations Several limitations of our current approach warrant acknowledgment and suggest directions for future research:

- **Computational Resource Constraints:** As noted in our methodology, the resource requirements of generating responses from multiple synthetic personas limited our ability to scale to larger question sets. Each synthetic persona requires a separate LLM instance, creating significant computational overhead for large-scale implementations. This constraint influenced our experimental design, potentially limiting the complexity of response patterns that could be captured and optimized.

- **Practical Implementation Barriers:** Beyond computational resources, implementing this approach faces challenges related to maintaining consistency across multiple model instances, managing prompt engineering complexity, and establishing verification methods for synthetic responses. Organizations adopting this technology would need specialized expertise and infrastructure to effectively deploy and maintain such systems.
- **Demographic Representation:** While our synthetic personas incorporated multiple demographic attributes, they necessarily represent simplified approximations of the complex intersectional identities that characterize real human diversity. Cultural contexts, life experiences, and situational factors that influence human responses were not fully captured in our current implementation.
- **Model-Specific Biases and Hallucinations:** The underlying LLM architecture inherently incorporates biases present in its training data, which may disproportionately affect the representativeness of certain demographic groups. Additionally, LLMs can produce hallucinated responses—generating plausible but factually incorrect information—which poses reliability challenges when simulating demographic responses to factual questions. These issues could propagate through our synthetic personas, potentially reducing accuracy for underrepresented populations or when addressing knowledge-dependent topics.
- **Question Domain Limitations:** Our experiments focused primarily on questions related to technology usage and preferences, which may exhibit different demographic response patterns than other domains such as health behaviors, political attitudes, or consumer preferences. The generalizability of our findings across diverse question types requires further investigation.

Future Research Directions Based on our findings and acknowledged limitations, we identify several promising directions for future research:

- **Multimodal Integration:** Expanding synthetic personas to incorporate non-textual responses, including visual preferences, voice patterns, or interactive behaviors could enhance their representational fidelity. This integration would more closely approximate the multimodal nature of human survey participation.
- **Dynamic Persona Adaptation:** Developing personas capable of evolving responses based on changing contexts or new information would better reflect the adaptive nature of human response patterns. This approach could incorporate temporal factors and life-stage transitions that influence demographic response characteristics.
- **Cross-cultural Validation:** Extending our methodology to diverse cultural contexts would test the generalizability of our approach and identify potential cultural variations in the relationship between demographic attributes and response patterns.
- **Explicit Demographic Optimization:** Future implementations could incorporate demographic distribution alignment directly into the fitness function, potentially yielding synthetic persona groups with even greater representational accuracy while maintaining response fidelity.
- **Computational Efficiency Enhancements:** Developing more efficient implementations that reduce the computational overhead of maintaining multiple synthetic personas would enable scaling to larger and more diverse question sets, potentially capturing more nuanced response patterns.

These future directions aim to address current limitations while expanding the potential applications of synthetic personas in demographic research, market analysis, and personalized content development. By combining the generative capabilities of large language models with the optimizing power of genetic algorithms, synthetic personas offer a promising approach to understanding and representing diverse demographic perspectives at scale.

6 Future Work

While our research demonstrates the effectiveness of synthetic personas for demographic response simulation, several limitations warrant further investigation. Our current implementation faces computational scaling constraints, the demographic attributes represent simplified approximations of complex identities, and the underlying language models incorporate biases that may affect representativeness. Addressing these limitations opens several promising research directions.

The most immediate opportunity for extending our work lies in enhancing the representational capabilities of synthetic personas. **Multimodal integration** would expand beyond textual feedback to incorporate visual inputs and outputs, enabling more comprehensive assessment of products with significant visual components. Extending synthetic personas to interpret user-provided images and generate visualizations or preference maps would more closely approximate the richness of human communication. A critical challenge in this area involves maintaining demographic consistency across modalities, requiring new evaluation metrics and validation approaches. Simultaneously, **advanced representational approaches** could improve both efficiency and fidelity. Dense vector embeddings could capture complex relationships between demographic attributes more effectively than discrete assignments, enabling more nuanced modeling of intersectional identities. Techniques such as variational autoencoders could facilitate exploration of persona "latent space," potentially revealing unexpected relationships between demographic characteristics and response patterns.

Methodological refinements offer another avenue for enhancing synthetic persona performance. While genetic algorithms proved effective in our implementation, **alternative optimization approaches** warrant exploration. Gradient-based methods might offer more efficient optimization for continuous persona representations, while reinforcement learning could enable more dynamic adaptation to changing question contexts. For computationally intensive evaluations, Bayesian optimization could reduce the function evaluations required to identify optimal configurations. Extending our framework to explicitly balance multiple objectives—response accuracy, demographic representativeness, and perspective diversity—could yield more versatile synthetic persona collections. Beyond optimization, **deeper analysis of trait effects** would enhance our understanding of how demographic characteristics influence responses. Factorial analysis of trait interactions could reveal complex conditional relationships, while causal models would improve both theoretical understanding and practical applications. Quantifying the relative importance of different attributes across question domains would enable more focused persona design.

A significant limitation of our current study is its focus on a single survey domain and a limited question set. Future research should investigate **cross-domain generalization capabilities** by testing whether personas optimized for technology usage transfer effectively to domains like healthcare preferences or consumer behavior, and by evaluating performance across a substantially larger and more diverse set of questions. Developing techniques for rapid adaptation to new domains with minimal additional optimization would enhance practical utility. Identifying persona characteristics that consistently influence responses across domains could lead to more robust implementations, while evaluating multi-domain training approaches could establish whether broader exposure produces more generalizable personas.

The potential to infer demographic characteristics from response patterns raises important **ethical considerations** regarding privacy and consent. As synthetic persona technologies advance, establishing comprehensive ethical guidelines becomes essential for responsible adoption. These should address transparency in synthetic persona deployment, appropriate limitations on demographic inference applications, and continuous monitoring for potential biases. Developing methodologies that effectively combine synthetic and human responses could leverage the strengths of both approaches while mitigating their respective limitations.

Looking forward, we envision synthetic personas evolving into dynamic representations capable of modeling demographic changes over time, adapting to diverse cultural contexts, and providing increasingly realistic simulations of human diversity. These advancements will position synthetic personas as a valuable complement to traditional research methods, enhancing our understanding of demographic perspectives at scale while maintaining ethical responsibility.

7 Conclusion

This research has demonstrated that integrating large language models with genetic algorithms creates an effective framework for generating synthetic personas capable of emulating demographic-specific response patterns. Our experimental findings establish that optimized synthetic personas can produce feedback closely aligned with real-world demographic distributions while generalizing effectively to novel questions.

Our three experiments yielded several significant findings. First, synthetic personas demonstrated the ability to produce responses that systematically vary with assigned demographic attributes, with age-based variations in technology usage patterns closely mirroring documented real-world trends. Second, genetic algorithm optimization substantially improved response accuracy, increasing training set alignment from 60.4% to 78.5% and test set accuracy from 62.6% to 68.8%, confirming effective generalization to unseen questions. Third, GA-optimized personas exhibited improved alignment with actual demographic distributions even for attributes not explicitly included in the optimization function, suggesting the emergence of realistic demographic representations through response pattern optimization. These results collectively support our fundamental hypothesis that synthetic personas, when properly constructed and optimized, can generate feedback that approximates genuine human responses across demographic groups.

This work advances the field of synthetic response generation in several important ways. We have developed a structured framework that integrates large language models with evolutionary optimization techniques, demonstrating the effectiveness of this combination for demographic-specific response generation. This methodological approach bridges previously separate research streams in natural language processing and optimization. Our experimental results provide empirical evidence that synthetic personas can maintain demographic consistency across diverse question types and contexts, confirming the feasibility of demographic-specific response simulation through appropriate model conditioning. Furthermore, we have shown that evolutionary optimization techniques can significantly enhance the demographic representativeness of synthetic personas through iterative selection based on response accuracy, establishing a novel approach to generating demographically realistic synthetic data. These contributions extend beyond technical implementation details to address fundamental questions about the nature of demographic response patterns and the capacity of language models to emulate these patterns when appropriately conditioned and optimized.

The demonstrated capabilities of optimized synthetic personas have significant implications for several applied domains. Our approach enables rapid, cost-effective simulation of demographic-specific feedback for product concepts, marketing materials, and service offerings without requiring extensive human participant recruitment. This capability can substantially accelerate the formative research phase of product development while reducing associated costs. Synthetic personas can identify potential demographic biases, ambiguities, or sensitivities in survey instruments before deployment, improving instrument quality and reducing the need for extensive pilot testing with human participants. The observed alignment between optimized synthetic personas and actual demographic distributions suggests applications in demographic modeling and forecasting, potentially enabling researchers to simulate demographic shifts and their associated response pattern changes. Understanding demographic-specific response patterns can inform more effective content personalization strategies, particularly for applications where explicit demographic data collection is limited by privacy considerations.

Importantly, we view these applications as complementary to, rather than replacements for, traditional human-centered research methods. The value of synthetic personas lies in their ability to extend research capabilities, increase efficiency, and identify potential areas for more focused investigation with human participants. Our findings demonstrate that the combination of large language models and genetic algorithms offers a promising approach for demographic research that addresses key limitations of traditional methods. By enabling more efficient, scalable, and adaptable investigation of demographic response patterns, synthetic personas represent an important methodological advancement in understanding human diversity at scale.

References

1. M. B. Akbar, L. Foote, C. Soraghan, R. Millard, and F. Spotswood, “What causes social marketing programs to fail? A qualitative study,” *Social Marketing Quarterly*, vol. 27, no. 2, pp. 99–116, 2021, iSBN: 1524-5004 Publisher: SAGE Publications Sage CA: Los Angeles, CA.
2. S. Wagner, D. Mendez, M. Felderer, D. Graziotin, and M. Kalinowski, “Challenges in Survey Research,” in *Contemporary Empirical Methods in Software Engineering*, M. Felderer and G. H. Travassos, Eds. Cham: Springer International Publishing, 2020, pp. 93–125. [Online]. Available: https://doi.org/10.1007/978-3-030-32489-6_4
3. M. Sahami, S. Dumais, D. Heckerman, and E. Horvitz, “A Bayesian approach to filtering junk e-mail,” in *Learning for Text Categorization: Papers from the 1998 workshop*, vol. 62. Citeseer, 1998, pp. 98–105.
4. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.
5. A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving language understanding by generative pre-training,” 2018, publisher: San Francisco, CA, USA.
6. P. Von Hilgers and A. N. Langville, “The five greatest applications of Markov chains,” in *Proceedings of the Markov Anniversary meeting.* Boston Press Boston, MA, 2006, pp. 155–158.
7. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *arXiv preprint arXiv:1412.3555*, 2014.
8. R. Pascanu, T. Mikolov, and Y. Bengio, “On the difficulty of training recurrent neural networks,” in *International conference on machine learning.* Pmlr, 2013, pp. 1310–1318.
9. A. Graves and A. Graves, “Long short-term memory,” *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012, iSBN: 3642247962 Publisher: Springer.
10. I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” *Advances in neural information processing systems*, vol. 27, 2014.

11. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
12. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, and A. Askell, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
13. A. Tamkin, M. Brundage, J. Clark, and D. Ganguli, "Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models," *arXiv preprint arXiv:2102.02503*, 2021.
14. J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. Leoni Aleman, D. Almeida, J. Altenschmidt, S. Altman, and S. Anadkat, "GPT-4 Technical Report," *arXiv e-prints*, p. arXiv: 2303.08774, 2023.
15. H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, and F. Azhar, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.
16. T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa, "Large language models are zero-shot reasoners," *Advances in neural information processing systems*, vol. 35, pp. 22 199–22 213, 2022.
17. H. W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, Y. Li, X. Wang, M. Dehghani, and S. Brahma, "Scaling instruction-finetuned language models," *Journal of Machine Learning Research*, vol. 25, no. 70, pp. 1–53, 2024, iSBN: 1533-7928.
18. Y. Fu, H. Peng, A. Sabharwal, P. Clark, and T. Khot, "Complexity-based prompting for multi-step reasoning," in *The Eleventh International Conference on Learning Representations*, 2022.
19. X. Wang, J. Wei, D. Schuurmans, Q. V. Le, E. H. Chi, S. Narang, A. Chowdhery, and D. Zhou, "Self-Consistency Improves Chain of Thought Reasoning in Language Models," in *The Eleventh International Conference on Learning Representations*, 2023.
20. R. Tourangeau, L. J. Rips, and K. Rasinski, "The psychology of survey response," 2000, iSBN: 0521576296 Publisher: Cambridge University Press.
21. N. R. Council, *Nonresponse in social science surveys: A research agenda*, 2013.
22. R. M. Groves, F. J. Fowler Jr, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, *Survey methodology.* John Wiley & Sons, 2011.
23. O. P. John and S. Srivastava, "The Big-Five trait taxonomy: History, measurement, and theoretical perspectives," 1999, publisher: University of California Berkeley.
24. N. Selwyn, S. Gorard, J. Furlong, and L. Madden, "Older adults' use of information and communications technology in everyday life," *Ageing & Society*, vol. 23, no. 5, pp. 561–582, 2003, iSBN: 1469-1779 Publisher: Cambridge University Press.
25. Statistisk sentralbyrå, "08411: Inntektsintervall, etter kjønn. Antall bosatte personer 17 år og eldre 1993 – 2023 [Income Interval by Gender: Number of Residents Aged 17 and Older, 1993–2023]." [Online]. Available: https://www.ssb.no/statbank/table/08411/
26. J. Sullivan, "The primacy effect in impression formation: Some replications and extensions," *Social Psychological and Personality Science*, vol. 10, no. 4, pp. 432–439, 2019, iSBN: 1948-5506 Publisher: Sage Publications Sage CA: Los Angeles, CA.
27. M. Grundetjern, P.-A. Andersen, M. Goodwin, and K. A. Borgersen, "NER Explainability Framework: Utilizing LIME to Enhance Clarity and Robustness in Named Entity Recognition," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence.* Springer, 2024, pp. 3–15.

Authors

Morten Goodwin is a professor of artificial intelligence at the University of Agder. He co-founded and is the lead researcher at AI Experts and serves as the deputy head at the Center for Artificial Intelligence Research (CAIR). With 20 years of academic experience, a Ph.D., and more than 200 peer-reviewed papers, his research focuses on deep neural networks. He received the NORA Award for his outstanding contribution to the AI community and is involved in numerous research projects spanning fundamental and applied AI research across multiple domains. He is a government-appointed member of both the

Privacy Board (Personvernnemnda) and the Technology Council (Teknologirådet).

Per-Arne Andersen received his Bachelor's and Master's degrees in Information and Communication Technology from the University of Agder, Norway. He also holds a Ph.D. from the same institution, where his doctoral research focused on reinforcement learning and its applications in industrial settings. Andersen is currently an Associate Professor at the University of Agder and a key researcher at the Centre for Artificial Intelligence Research (CAIR). His research interests encompass various domains of artificial intelligence, including machine learning, deep learning, reinforcement learning, and cybersecurity.

Morten Grundetjern is currently pursuing his PhD in machine learning at the University of Agder. He holds a bachelor's degree as a data engineer and a master's in ICT focusing on AI. Before his academic career, he worked as an electrician, completed military service in Lillehammer, and attended folk high school in Elverum. He previously worked at Sikri where he was involved in various projects. His research is conducted as part of CAIR (Center for Artificial Intelligence Research) at the University of Agder.